Contents lists available at ScienceDirect

The Leadership Quarterly

journal homepage: www.elsevier.com/locate/leaqua

Where is “behavior” in organizational behavior? A call for a revolution in leadership research and beyond

George C. Banks[a,*], Haley M. Woznyj[b], Claire A. Mansfield[a]

[a] University of North Carolina, Charlotte, United States
[b] Longwood University, United States

ARTICLE INFO

Keywords:
Behavior
Leadership
Organizational behavior
Questionnaires
Endogeneity bias

ABSTRACT

Behaviors can be characterized as “the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli.” (Levitis et al., 2009). The study of behavior is a critical component of theory advancement in the area of leadership. Yet, a large number of leadership studies conflate behavioral and nonbehavioral concepts. First, our manuscript offers a theoretical discussion of why the absence of research on behavior is a growing concern for the advancement of theory in leadership. Evidence from a systematic review (k = 214) indicates that of 2338 variables only 3% are behavioral in nature (19% of studies include at least one behavioral measure). Second, we present a framework of behavior to better distinguish leader (follower) behaviors from other concepts. Finally, we provide a set of methodological recommendations to ensure alignment between theoretical conceptualizations and methodological choices.

Behaviors can be characterized as “the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes.” (Levitis et al., 2009; p. 103). The field of leadership and organizational behavior more broadly aims to build and test theories around behavioral phenomena (Andersson et al., 2013; Ashkanasy, 2013; Liu et al., 2019). Entire sub-fields have emerged around the study of leader behaviors (for an empirical review see Banks et al., 2018a) as well as critical behavioral outcomes including task performance (Motowidlo & Van Scotter, 1994), citizenship behaviors (LePine et al., 2002), and counterproductive work behaviors (Dalal, 2005). However, despite the importance of leader behaviors and follower behaviors as outcomes, many have argued that the study of behaviors has become rather passé, with subsequent consequences for theoretical advancement (Baumeister et al., 2007).

In the broader field of organizational behavior, it is not uncommon to use psychological variables as proxies of behaviors, such as turnover intentions to represent turnover (Tett & Meyer, 1993) and applicant attraction to represent job acceptance (Chapman et al., 2005). In leadership research, evaluations of leaders (y_1; measured via questionnaires) are commonly used to indicate leadership behaviors (x_1; Antonakis et al., 2016; Day, 2014; Fischer et al., 2020). Similarly, scholars who seek to investigate outcomes of leadership often study counterproductive work *behaviors* as well as organizational citizenship *behaviors*, but actually capture self- and other-evaluations of behavior via questionnaires (Berry et al., 2012; LePine et al., 2002). Advancing scholarship around perceptions of behaviors (i.e., awareness through senses of behaviors) and evaluations (i.e., the appraisal or assessment of behavior) is important for theory and practice (Chan, 2009).

However, using either perceptions or evaluations as proxies for leader or follower behaviors can be problematic for three different reasons ($x_1 \neq y_1$). First, the overuse of proxies leads to weak theory testing in which perceptions and evaluations are used to represent actual behaviors, leading to an inaccurate (or at the least, misleading) accumulation of knowledge. This begs the question: are evaluations of leaders (and followers) capturing more than just “do I like my boss?” (Yammarino et al., 2020). Second, because of the overreliance on questionnaires of perceptions and evaluations, we often have a limited ability to make causal inferences ($x_1 \rightarrow y_1$ vs. $x_1 \neq y_1$). Third, there is a potential for unique types of bias to be introduced (Fischer et al., 2020). Meta-analytic evidence indicates that supervisor-rated task performance is prone to bias as a result of poor reliability and questionable validity (Rothstein, 1990; Viswesvaran et al., 1996). Bias in evaluations of leaders can result in discrimination which may lead,

* Corresponding author at: University of North Carolina at Charlotte, Belk College of Business, 9201 University City Blvd, Charlotte, NC 28223, United States.
E-mail address: gcbanks@gmail.com (G.C. Banks).

https://doi.org/10.1016/j.leaqua.2021.101581
Received 24 February 2021; Revised 29 September 2021; Accepted 2 October 2021
Available online xxxx
1048-9843/© 2021 Elsevier Inc. All rights reserved.

for instance, to a lack of women in upper echelons of organizations (Martell et al., 1996; Samuelson et al., 2019).

These limitations of perceptions and evaluations are highly interconnected as well; unique biases can impede causal inference, which can decrease the quality of theory testing. Again, this is not to say that perceptions and evaluations do not matter. Rather, perceptions and evaluations are simply not leader (or follower) behaviors and should not be treated as such in our theories. This then raises the question, if the study of perceptions and evaluations has become so dominant, what has happened to the study of human behavior in research on leadership and other topics? If past behavior is the best predictor of future behavior (Mumford & Owens, 1987; Owens, 1976; Owens & Schoenfeldt, 1979), the domain of leadership and organizational behavior, generally speaking, should make more of an effort to study it.

The current manuscript is organized into three major sections. First, we review various definitions of behavior and explain why the absence of behavior in theories limits our ability to build and test theories. We also review the growing calls of concerns from scholars regarding the lack of behavioral studies in leadership and, more broadly, areas of management and applied psychology. We present evidence from a systematic review of leadership and organizational behavior studies to illustrate the extent to which behaviors are studied to highlight the full extent of the issue. Second, we introduce a new framework of behavior which serves to stimulate thinking around the role of behavior in leadership theories. Finally, we present a set of action-based methodological recommendations to guide future leadership research and organizational behavior scholarship, broadly speaking. While we focus the majority of the current review on the domain of leadership and organizational behavior, the key implications are made relevant for other areas of the organizational sciences, such as strategic management, entrepreneurship, political science, and international business where behaviors play an important role in theories.

The absence of "behavior" impedes theory advancement

Before one can understand how the absence of leader and follower behavior impedes theory advancement, it is necessary to clearly conceptualize behaviors. A cursory review of organizational behavior textbooks commonly used in undergraduate and graduate classes shows that these texts almost always define organizational behavior as a field, but not behaviors themselves[1]. The omission of the definition of behavior in organizational behavior is not unique to the textbooks in this area (Levitis et al., 2009). To help create clarity, we provide a list of illustrative definitions of behaviors from a variety of different types of sources in Table 1. These examples demonstrate a range of possible conceptualizations although the list is by no means exhaustive. The sources range from popular mainstream dictionaries, such as Merriam Webster who define behavior as "anything that an organism does involving action and response to stimulation" to definitions published in more traditional academic works. For instance, Dretske (1988) intentionally oversimplified behavior and stated that all behavior involves "some kind of bodily movement, and of each such movement as having some more or less unique cause" (p. 1).

Perhaps one of the most thorough studies of behavioral definitions was conducted by Levitis et al. (2009), who sought to understand how behavior was defined and how it should be defined in the academic literature. These authors completed a systematic review of published definitions across a variety of disciplines. They then conducted a survey of 174 subject matter experts from multiple disciplines and uncovered a lack of consensus regarding how to conceptualize behavior. Example survey prompts included "a person decides not to do anything tomorrow if it rains," "a person sweats in response to hot air," and "flocks of geese fly in V formations." The authors attempted to rectify the lack of consensus and as previously stated, ultimately defined behavior as "the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes." (p. 103). Based on this work, we put forth that the definition by Levitis et al. (2009) is the most rigorous published and adopt it in the current work.

Table 1
Example definitions of behavior.

Source	Definition examples
Dictionary sources	
1. Merriam-Webster	▪ "Anything that an organism does involving action and response to stimulation"
2. Dictionary.com	▪ "The aggregate of responses to internal and external stimuli"
Academic literature	
3. Ajzen and Fishbein (1977; p. 889)	▪ "Behavioral criteria consist of one or more observable actions performed by the individual and recorded in some way by the investigator"
4. Dretske (1988; p. 1)	▪ "A behavior involves some kind of bodily movement, and of each such movement as having some more or less unique cause"
5. Levitis et al. (2009; p. 103)	▪ "Behaviour is the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes"
6. Henriques and Michalski (2020)	▪ Movements that generate measurable effects

Although we focus primarily on behaviors in this manuscript, we compare and contrast behaviors to perceptions and evaluations. As previously mentioned, we characterize perceptions in this context as the awareness through one's senses of behaviors and evaluations as the appraisal of behavior. We do not explicitly focus on other psychological states, such as attitudes, feelings, or behavioral intentions, in order to simplify the discussion. However, these psychological states are also important and can similarly be theoretically conflated (Fischer et al., 2020).

In recent years, some areas of leadership have made attempts to theoretically and empirically distinguish behaviors, perceptions, and evaluations. For example, Antonakis et al. (2016) have defined charismatic leadership behaviors as "values-based, symbolic, and emotion-laden leader signaling." Similarly, Banks et al. (2021a) recently defined ethical leadership behaviors as "signaling behavior by the leader (individual) targeted at stakeholders (e.g., an individual follower, group of followers, or clients) comprising the enactment of prosocial values combined with expressions of moral emotions." These works emphasized distinguishing between specific leader behaviors and follower evaluations.

Theoretical specificity in the study of leader and follower behavior

As we continue our discussion, we distinguish between different types of concepts as well as how they are measured. Drawing upon Podsakoff et al. (2016), we define a concept as "cognitive symbols (or abstract terms) that specify the features, attributes, or characteristics of the phenomenon in the real or phenomenological world that they are meant to represent and that distinguish them from other related phenomena" (p. 161). It is important to distinguish between concepts and the measures used to assess those concepts (Arthur & Villado, 2008). For example, cognitive ability tests (the measure)

[1] We exclude the list of textbooks we examined so as not to unfairly single out textbook authors. We believe omitting the definition of behavior is common.

mostly measure cognitive ability (the concept). There is always the possibility of minimal contamination in cognitive ability tests, such that they also measure reading comprehension skills, for example, and there is deficiency in that they do not measure all aspects of cognitive ability. However, other approaches to assessing cognitive ability (the concept), such as situational judgement tests or structured interviews (alternative measures) may also assess personality and domain knowledge, among other things that may or may not be of interest, thereby resulting in substantially more contamination. Because of poor, imprecise, and/or proxy measures, our assessments of intended concepts (e.g., behaviors) may get conflated with other concepts. Consequently, there is a need to distinguish between theoretical concepts as well as the way in which those concepts are operationalized as variables.

To further illustrate this point, and to highlight how the absence of leader and follower behaviors is harmful to theory, we offer two illustrative examples, depicted in Figs. 1 and 2. The aim of these figures is to demonstrate the importance of theoretical specificity in the study of concepts, such as perceptions, evaluations, and behaviors. We begin with Fig. 1, which details a simple illustration of a follower working in a potato chip factory. This first example is intentionally simplified in order to provide clarity, and we later provide a more complex example that captures other work environments commonly studied in leadership research. In the figure, we offer a minimum of three different concepts measured seven different ways that could represent various conceptualizations of task performance as perceptions, evaluations, and behaviors. In reality, the measures depicted in Fig. 1 actually assess a variety of different concepts, highlighting the difference between measures and theoretically important concepts.

We first start with *follower performance behavior*, which can be measured using the number of potatoes peeled (Fig. 1, #1), the speed to complete the peeling of potatoes (Fig. 1, #2), and the numeric weight of potatoes (Fig. 1, #3), depending on the aspect of performance that is of interest. Next, there are *evaluations of follower performance,* measured using behaviorally anchored ratings of performance, as evaluated by the leader (Fig. 1, #4). Structured evaluations are prone to biases such as halo or horn effects (Robbins & Judge, 2008), among others, which cloud the follower's true standing on the concept due to leader's tendencies to rate their followers leniently or harshly. Leaders (often formal supervisors) can also evaluate the task performance of their followers in an unstructured way using qualitative descriptions of performance (Fig. 1, #5). Such ratings are also prone to the biases of the leader and how they perceive their reality (Aguinis & Solarino, 2019). Thus, the concept moves from the behavior of task performance and transforms into evaluations of task performance.

Next, there are *perceptions and/or evaluations of performance,* measured using a follower's self-report of performance (Fig. 1, #6). Self-report questionnaires are commonly used in organizational behavior studies and are a helpful tool to researchers, when used appropriately (Chan, 2009). However, researchers tend to over-rely on self-report questionnaires to assess a wide variety of concepts that self-report questionnaires are not best equipped to measure, such as behaviors (Antonakis, 2017; Fischer et al., 2020). The overreliance on self-report measures is problematic because there is evidence to suggest that self-report surveys have inherent flaws that make the assessment of behaviors difficult and conflate behaviors with perceptions and evaluations (Antonakis et al., 2016).

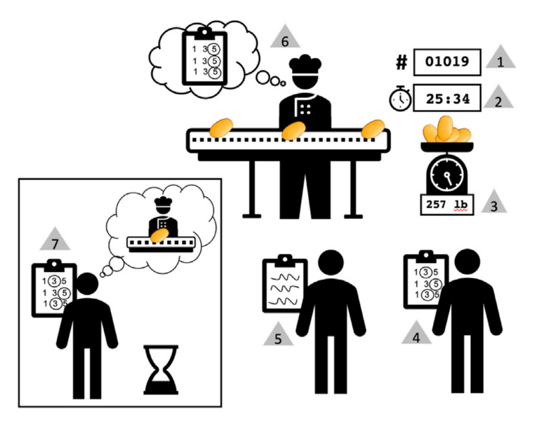
For instance, Carpenter et al. (2014) conducted a meta-analysis on the relationship between self-reported and other-reported organizational citizenship behaviors. They found that self- and observer-ratings are correlated at $\rho = 0.26$, which is low enough to suggest that they are not measuring the same concept. As another example, Furnham and Stringfield (1998) examined 360-degree feedback ratings and found that there was little congruence between follower self-ratings of performance and ratings from peers, supervisors, and consultants. A majority of followers perceive their performance behavior to be above average, which is not possible (Meyer, 1975). As a final example, Liao et al. (2009) found differences in perspectives between leaders and followers and, moreover, differences among peers of different levels of status when evaluating organizational activities. These studies suggested that even when different people experience and/or witness the same situation, they may perceive and evaluate that situation very differently. The difference in concepts is theoretically interesting and important to recognize.

Evidence from the social desirability literature also suggests that followers may not be aware of or may not accurately know themselves (Paulhus, 1984). For instance, evidence from the faking literature suggests that participants tend to respond to self-report surveys with answers that are untrue of themselves, or may respond in socially desirable ways (Nederhof, 1985). Faking has been expressed as a concern for tests such as situational judgement tests (Nguyen et al., 2005) and personality tests (Goffin & Christiansen, 2003; LeBreton et al., 2007). As additional evidence, respondents' memory recall of events that have happened in the past tends to be flawed. Johns and Miraglia (2014) meta-analytically explored the comparison of employees' self-reports of absenteeism and organizational records. The authors found that, when asked to self-report their absences, employees tended to underreport the number of days they missed work. The underreporting was heightened for general absences, as compared to absences due to sicknesses. Thus, self-reports of behaviors can be inherently flawed measures of objective behavior.

Finally, for this first example depicted in Fig. 1, there is again *evaluation of followers' performance* (Fig. 1, #7). This is measured using the memory of a leader of a follower's performance from some point in time in the past (e.g., over a 6-month or 12-month time period). This is arguably the most common type of task performance measure in leadership research and is widely used in practice via annual evaluations. Similar to self-rated evaluations of performance, these measures are subject to memory recall biases and can suffer from primacy and recency effects (see Steiner & Rain, 1989), whereby only select instances of behaviors are considered and reported. Leaders are also under pressure from followers to give positive evaluations, which restricts the reported variance of the measured concept (Meyer, 1975). It is worth noting that when two or more supervisors are used to complete global evaluations of performance, meta-analytic evidence indicates that inter-rater reliability is quite low (Rothstein, 1990; Viswesvaran et al., 1996). Thus, recalled global evaluation of follower performance again is contaminated with (or deficient of) extraneous factors that make it an 'impure' assessment of behavior.

Moving to the second example (see Fig. 2), we present a more complicated illustration of a collection of related, but distinct, concepts of leader activities. Our first example focused on follower behaviors and here we switch the focus to leader behaviors. First, there is the concept of organizational policies (Fig. 2, #1) which may be used as a proxy for leader signaling, especially if the leader states such policies in a meeting or email. This is typically measured by qualitatively and/or quantitatively coding policies written by the organization. There are a number of potential sources of systematic and random error introduced here, as well as contamination and deficiency in the validity of such measures. In this case, the measurement error is contingent upon the specific measure and analysis technique (e.g., content analysis, Qualitative Content Analysis-QCA; Natural Language Processing-NLP). The validity may also be contingent upon the specific measure, background assumptions, and experiences of the coder(s). In addition, although policies might be formalized in an employee handbook, this does not necessitate that the leader implemented them or that they are implemented in a similar way across leaders (Ostroff & Bowen, 2016). Thus, relying on these as a measure of leadership activities also come with limitations.

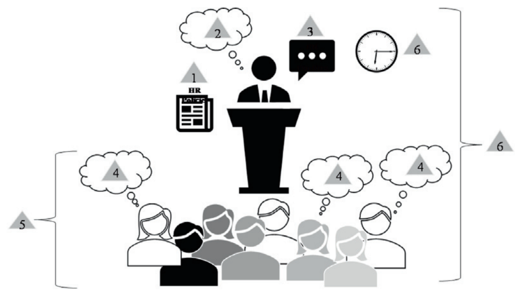
Second, there are leaders' *self-evaluations of behavior* (Fig. 2, #2) (Hammer et al., 2009). Here, the supervisor completes a self-rating of behavior using a validated questionnaire. Reliability of the measure



#	Theoretically intended concept	Variable measure	Potential sources of variation Reliability *Systematic & Random Error*	 Validity *Contamination & Deficiency*
1	Follower performance behavior	Number of potatoes peeled	N/A	N/A
2	Follower performance behavior	Speed to complete defined task	Accuracy of the clock	N/A
3	Follower performance behavior	Numeric weight of potatoes	Accuracy of the scale	N/A
4	Evaluation of follower performance	Behaviorally anchored ratings of performance, as evaluated by the supervisor	Idiosyncratic interpretation of items, mood, time of day	Self-fulfilling prophecies; halo/horn bias effects; PA/NA
5	Evaluation of follower performance	Qualitative description of performance, as evaluated by supervisor	Mood, personal experiences and background, time of day	Background, assumptions and experiences of the evaluator
6	Self-perceptions of follower performance	Self-evaluation of performance	Idiosyncratic interpretation of items, mood, time of day	Self-awareness; need for achievement; core self evaluation; conscientiousness, narcissism
7	Evaluation of follower performance	Memory of performance from some point in time in the past	Idiosyncratic interpretation of items, mood, time of day	Memory, primacy/ recency effects, relative comparisons

Note: This list is not exhaustive–there are other measures of performance; N/A indicates that this source of variation is not likely to occur, though it could be possible. PA/NA = Positive Affectivity/Negative Affectivity.

Fig. 1. The depiction of concepts related to follower task performance.



#	Theoretically intended concept	Variable Measure	Potential sources of variation Reliability *Systematic & Random Error*	 Validity *Contamination & Deficiency*
1	Organizational policies	Qualitatively and/or quantitatively coded policies written by the organization	Contingent upon specific measure and analysis technique (e.g., content analysis, QCA, NLP)	Contingent upon specific measure; background, assumptions, and experiences of the coder
2	Self-perceptions of leader behavior	Leader's self-rating on a questionnaire	Idiosyncratic interpretation of items, mood, time of day	Self-awareness; need for achievement; core self evaluation; conscientiousness, narcissism; familial status
3	Leader behaviors	Audio and/or video recording of a leader's speech	Contingent upon specific measure and analysis technique (e.g., QCA, NLP)	Background, experiences, management training, familial status, societal culture, org. culture, org. expectations
4	Follower evaluations of a leader's behavior	Individual-level evaluation of a leader	Idiosyncratic interpretation of items, mood, time of day	Selective perception, familial status, home demands, societal culture, psychological contract, relative comparison to others
5	Team evaluations of a leader	Aggregated evaluations of a leader	Idiosyncratic interpretation of items, mood, time of day	Prone to the aggregate concerns of individuals
6a	Leadership culture	Objective measures of time, such as hours worked, start and end time of meetings, workdays etc.	Contingent upon specific measure	N/A
6b		Aggregation of subjective and/or objective measures of a leader	Contingent upon specific measure	Industry, societal norms

Note: QCA = Qualitative content analysis; Natural Language Process = NLP

Fig. 2. The depiction of concepts related to leader behavior.

may be influenced by idiosyncratic interpretation of items, mood, and time of day. Validity could be affected by self-awareness, need for achievement, core self-evaluations, as well as individual differences, such as conscientiousness and narcissism. Third, *leader behaviors* could be measured using an audio and/or a video recording of a leader's speech (Fig. 2, #3). The reliability of this measure might be contingent again upon the measure and analysis technique (e.g., QCA; NLP). Validity could be influenced by background experiences, management training, societal culture, as well as organizational culture and expectations. The validity here would be based on the extent to which the score extracted from the speech (if done by coders or algorithms) accurately reflects the level of family-supportive HR practices in the speech.

Fourth, another potential concept is follower *evaluations of leader behavior* (Fig. 2, #4). Questionnaires are arguably one of the most common measures used in leadership research (Fischer et al., 2020). Once again, reliability is affected by idiosyncratic interpretation of items, mood, and time of day. Validity could be influenced by selective perception, gender of the leader (e.g., women are evaluated differently than men), societal culture, and/or relative comparison to others. Fifth, there are team *evaluations of a leader behavior* (Fig. 2, #5). The concept could be measured using individual follower evaluations of a leader via a questionnaire that are then aggregated to the team level of analysis. There are some similar, but also unique reliability and validity concerns at the team level of analysis, compared to individual follower evaluations.

Finally, one may measure the concept of *leadership culture* (see Fig. 2, #6a and #6b), or shared perceptions of leader behaviors in the team or organization. The variables measured might be objective measures of time, such as hours worked or start and end times of meetings as well as workdays. Alternatively, the culture (or climate) measure might be the aggregation of subjective and/or objective measures of the leader. Reliability of the measure would be contingent upon the measure itself. Contamination and deficiency are likely less of a concern for the objective measure of leadership culture and are likely very relevant concerns for the perceptual measure of culture. Of course, there are many other concepts that could be measured in this scenario. For instance, of potential importance could be leader's behavioral *intentions*. That is, the leader intended to allow their followers to leave on time, but due to a looming deadline, the followers were required to stay late. While the intention (leaving on time) is very different from the actual behavior (staying late), there may be instances where that is of theoretical interest.

What would be different in leadership research with the study of behavior? There are certainly times where a subordinate's subjective evaluation of leader behavior would be consistent with objective behavioral measures. That is to say, one would expect there to be a moderate to large magnitude correlation between the objective measure of leader behavior and followers' evaluation of the behavior. Yet, there can be both contamination and deficiency in follower evaluations of leaders. Consider a simulation by Martell et al. (1996) which showed that five percent of variance in evaluations being driven by one's gender can result in the absence of women at upper echelons of organizations. The same overall result also occurred when only one percent of bias existed based on gender (for a more recent and robust simulation see Samuelson et al., 2019). Bias can stem from subjective evaluations or objective measures of behavior that represent systemic inequality. Hence, even small amounts of bias can be important contributing factors in leadership research. Thus, using behaviors can help to create a more accurate knowledge base.

As another example of what would be different if we, as a field, actually studied behaviors is that we would know what behaviors leaders engage in to *cause* social influence. For instance, Banks et al. (2021a; 2021b) reconceptualized ethical leadership behaviors and conducted a series of studies that (1) identified eight behaviors and (2) showed that the behaviors caused improved leader evaluations and follower performance as well as a reduction in counterproductive behaviors. This latter work also produced a machine learning algorithm that allowed for the automatic scoring of text (e.g., emails, meeting transcripts; speeches). What are the implications of this work? We now have a set of behaviors on which to formally train leaders to engage in more ethical signaling. Leaders can also receive more objective feedback on their behaviors. This work mirrors work that has already been completed by Antonakis et al. on charismatic leadership (2011; 2012; 2016). Historically charisma was thought to be something leaders were born with and that it could not be trained. Antonakis et al. demonstrated that through the use of 12 charismatic leadership tactics (CLTs) leaders could be trained to be charismatic and that such behaviors drive important outcomes, such as task performance.

In sum, the key point is that behavior and subsequent perceptions and evaluations are not the same concepts, nor are they interchangeable ($x_1 \neq y_1$), despite some research treating them as such. There may be times where scholars are theoretically interested in each of these concepts. However, problems can arise when perceptions and evaluations are used as proxies for behaviors. Consequently, it is important to keep these distinctions in mind when building and testing leadership theory.

Growing concerns regarding the lack of behavioral studies in leadership and organizational behavior

There is a growing concern in the area of leadership and the broader field of organizational behavior that the study of human behavior is disappearing from our scientific investigations (Fischer et al., 2020). To better understand the extent of such concerns, as well as their theoretical consequences, we summarize in Table 2 a review of arguments in the literature that have covered this matter. We provide quotes from leadership scholars (Antonakis et al., 2016; Banks et al., 2021a; Day, 2014; Fischer et al., 2017) but also the broader field of organizational behavior to highlight the full extent of these concerns.

For instance, Antonakis et al. (2016) stated "Currently, there seems to be a create-a-questionnaire bandwagon sweeping through our field to measure all sorts of constructs; many of these are poorly defined and operationalized… It certainly is a "quick and dirty" way to obtain data. However, such measures do not get at what charisma [behavior] is—in terms of an independent variable—no matter how big the statistical hammers, used to "confirm" the factor structure of the measures are." (p. 306). As a result of this overreliance on self-reports or observer evaluations instead of objective measures of behavior, there may be a misalignment between theory that is being tested and developed and the methodology being used.

As another example, Baumeister et al. (2007) highlighted the absence of behavioral evidence in the broad social science literature theories (p. 397):

> *Some psychological subdisciplines have never directly studied behavior, and studies on behavior are dwindling rapidly in other subdisciplines … Behavioral science today…mostly involves asking people to report on their thoughts, feelings, memories, and attitudes. Occasionally they are asked to report on recent or hypothetical actions. Or, somewhat differently (and more rarely), reaction times, implicit associations, or memory recall might be assessed in the service of illuminating a cognitive process. But that is as close as most research gets. Direct observation of meaningful behavior is apparently passé.*

There are two critical takeaways from reading Table 2. First, there is a concern from scholars in the field that self-report surveys are being over-relied on to study proxies of behaviors (Fischer et al., 2020). The frequency with which such concerns are appearing in the published literature seems to be growing. What is more, these concerns cross leadership as well as multiple subareas of organizational behavior (e.g., workplace deviance, performance). Perhaps the increase in reliance on self-report surveys stems from the fact that they are easy to administer (Day, 2014). Moreover, there is increased pressure to publish in today's fields of management and applied psychology which may resign scholars to prioritize data collection methods like surveys that are fast and easy (Aguinis et al., 2019). As we explained above in Figs. 1 and 2, the use of surveys is typically an inexact way to measure behaviors.

A second major takeaway is that there are consequences for leadership theories as a result of this problem. Of course, there are method-

Table 2
The absence of behavior impedes theory advancements in leadership and organizational behavior.

Author	Quote	Theoretical consequence
Spector (1994; p. 385)	Frequent discussions of the problems with self-reports can be found throughout the OB literature…Indeed self-reports have been used too frequently to address research questions that they are unable to adequately answer.	Misalignment between theory and methodology
Donaldson and Grant-Vallone (2002; p. 264)	Accurate measurement of organizational behavior is essential for advancing the field. Despite its importance, measurement in organizational settings is often referred to as one of the main shortcomings of organizational behavior research… [Self-report] measures are common because they are relatively easy to obtain and are often the only feasible way to assess constructs of interest.	Absence of behavioral evidence in our theories
Beal and Weiss (2003; p. 443)	Among these biases [of reports of past behavior] are greater influence of recent events on judgments of frequency of events, greater influence of salient events on event frequencies, identity or scenario-based biases of subjective experiences, and more general memory biases based on current mood states (e.g., mood congruence). Altogether, research indicates that people's summaries of their own states, experiences, and behaviors are poor reflections of the actual history of those states, experiences, and behaviors.	Misguided theory
Johnson and Turner (2003; p. 312)	Observation is an important method because people do not always do what they say they do.	Misguided theory
Bono and Judge (2004; p. 907)	As we are aware of no field studies that used behavioral (as opposed to perceptual) measures of transformational leadership, it is hard to know what effect using more rigorous measures might have had on the results.	Misguided theory
Adler, Thomas, and Castro (2005; p. 5)	Studies examining the validity of self-report suggest that evaluations of oneself may be biased by social desirability and have only a moderate relationship to objective assessments.	Misguided theory
Baumeister et al. (2007; p. 397)	Some psychological subdisciplines have never directly studied behavior, and studies on behavior are dwindling rapidly in other subdisciplines … Behavioral science today…mostly involves asking people to report on their thoughts, feelings, memories, and attitudes. Occasionally they are asked to report on recent or hypothetical actions. Or, somewhat differently (and more rarely), reaction times, implicit associations, or memory recall might be assessed in the service of illuminating a cognitive process. But that is as close as most research gets. Direct observation of meaningful behavior is apparently passé.	Absence of behavioral evidence in our theories
Stewart, Bing, Davison, Woehr, and McIntyre (2009; p. 208)	Heneman, Heneman, and Judge (Heneman, Heneman, & Judge, 1997) noted that employees may distort their responses on such measures to avoid describing them-selves in negative terms… Because self-reported answers can be difficult to verify, response distortion may also occur.	Misguided theory
Day (2014; p. 862)	Questionnaires remain a popular (if misguided) approach to studying leadership. If you design and publish a brief, easy-to-administer survey questionnaire, there is little doubt that researchers will use it. But we should not lose sight of the fact that the map is not the territory, and simply labeling a questionnaire as a measure of 'leadership' does not mean that it actually measures leadership.	Absence of behavioral evidence in our theories
Antonakis et al. (2016; p. 306)	Currently, there seems to be a create-a-questionnaire bandwagon sweeping through our field to measure all sorts of constructs; many of these are poorly defined and operationalized… It certainly is a "quick and dirty" way to obtain data. However, such measures do not get at what charisma is—in terms of an independent variable—no matter how big the statistical hammers, used to "confirm" the factor structure of the measures are.	Poorly defined concepts in organizational behavior theories
Fischer et al. (2017; p. 1736)	Retrospectivity, unspecified time frames, and person-whole approaches severely limit inferences about temporality and thereby causality.	Limited inferences of temporality and causality in theories
LeBaron, Jarzabkowski, Pratt, and Fetzer (2018; p. 241; p. 249)	People are not always cognizant of what they do and how they do it even in the moment of performance, and they may be even less aware in hindsight… While surveys and interviews are retrospective as they ask people to remember and interpret behavior, video recordings give researchers access to the details of real-time performance.	Misalignment between theory and methodology
Bostyn, Sevenhant, and Roets (2018; p. 1084–1085)	We argue that hypothetical-dilemma research, while valuable for understanding moral cognition, has little predictive value for actual behavior and that future studies should investigate actual moral behavior along with the hypothetical scenarios dominating the field … Accordingly, whether or not hypothetical moral judgment is related to real-life behavior is prone to become a matter of public interest.	Misalignment between theory and methodology
Banks et al. (2021a)	While this concern is not unique to the ethical leadership domain, it remains problematic that there continues to be a conflation between ethical leader behaviors and followers' evaluations of the leader's values, traits, and behaviors	Misalignment between theory and methodology

ological and analytic issues that are relevant to consider in terms of studying behavior. However, the absence of behavior in our research is a theoretical problem first and foremost. For example, it can lead to misguided theory. Because of the overreliance on surveys, literature areas are accumulating knowledge that builds results using measures that do not accurately reflect the concepts in question. Bono and Judge (2004) noted in the context of transformational leadership, "as we are aware of no field studies that used behavioral (as opposed to perceptual) measures of transformational leadership, it is hard to know what effect using more rigorous measures might have had on the results" (p. 907). Consequently, the absence of behavioral evidence may fundamentally alter our theoretical understanding of prominent concepts.

While the review of concerns presented in Table 2 is by no means exhaustive, these arguments point to the fact that there is a problem in the area of leadership and these points serve as exemplars of the theoretical consequences that must be addressed. However, the points are, to some extent, anecdotal. To determine the degree to which these concerns are founded, we systematically reviewed the literature for trends regarding the presence or absence of behavioral evidence in leadership and organizational behavior research.

Evidence of the absence of behavioral evidence in leadership and organizational behavior

Systematic search. In order to analyze the area of leadership and the broader field of organizational behavior for trends in the study of behaviors, we took two approaches. One approach included a focus on *The Leadership Quarterly* and *Journal of Organizational Behavior* (JOB), given the relevance of their mission for the study of leadership and organizational behavior. Consequently, we reviewed all articles published in 2019 from these journals. In the second approach, we conducted a broad systematic search by selecting several representative journals and identifying meta-analytic reviews published in those journals. Our aim with this alternative approach was to limit systematic bias in our review of primary studies. This method had several advantages: (1) by not restricting the search to specific journals in this approach, we are able to show the rate of study of behavior across areas of leadership and organizational behavior not limited to a specific set of journals or disciplines; (2) the selection of topics within meta-analytic reviews shows that the primary studies are in very popular and relevant areas (popular enough for a sufficient amount of data to exist for a meta-analytic review); (3) we randomly selected studies within these meta-analytic reviews to eliminate bias; (4) we achieved scale to demonstrate that the results of the review are not a reflection of random-sampling error; and finally, (5) we suggest that these studies are representative of the field as they were the input for the meta-analytic reviews in elite journals (meta-analytic reviews are the most highly cited type of studies in management and applied psychology according to Antonakis et al., 2014; Judge et al., 2007).

While consensus of which journals are the most representative is unlikely, we believe that we identified four journals that allow us to draw general conclusions about the study of behavior. Again, recall that our focus is not simply leader behaviors, but also common outcomes of interest to leadership scholarship (e.g., counterproductive work behaviors; organizational citizenship behaviors). These journals were *Journal of Management* (ranked #3 out of 217 Management journals in 2018 according to the 2019 Clarivate Analytics Journal Citation Reports), *Academy of Management Journal* (ranked #9), *Personnel Psychology* (ranked #12), and *Journal of Applied Psychology* (ranked #25). In 2019, we selected all meta-analytic reviews from 2017 and 2018 that were published in these journals. As our focus was on behavior, we included only meta-analytic reviews that focused on individual and team-level behaviors[2]. This left 33 meta-analytic reviews in total. We also excluded articles that used simulated data, but we later offer examples of ways simulations can be used to model behaviors.

We then used a random-number generator and selected three primary studies from each meta-analytic review. This procedure gave us 99 journal articles. If by the time we completed the coding, saturation had not been achieved, we then planned to continue to retrieve one randomly identified journal article per meta-analytic review and continue coding in rounds until saturation was reached. Saturation here is defined as the point that information observed becomes redundant and new insights are no longer emerging (Becker et al., 2008). The primary inference of interest in this context is the rate at which behavior is being studied. Thus, saturation was achieved when the general conclusion or theme (the rate that behavior is being studied) does not change with the addition of new data. We went through two iterations.

Coding. In each article, the key pieces of information coded were (1) study design (experimental, quasi-experimental, observational), (2) total number of variables described in the methods section, and (3) number of variables that measure actual behavior. To ensure the reliability of our findings, inter-rater reliability was calculated on a subset of articles between the second and third author who coded. Across 63 coding decisions, there was 90% agreement and a Cohen's kappa of 0.81 (Data for journal, year, study design, and aggregated variable count can be found here: https://osf.io/qh2yr/?view_only=8737b3b6d67a4837bf9f042bb61808c9). The second and third authors then divided up and independently coded the remaining articles.

To classify variables as behaviors, we drew upon the previously stated conceptualization from Levitis et al. (2009) we discussed to this point in the manuscript. When using this definition during our coding, we took a liberal approach. That is, if there were questions of whether a particular measure should be categorized as a behavior, we erred on the side of including it, but discussed this decision within the author team. While we were relatively lenient in our inclusion criteria, this allows for a richer, more diverse array of behaviors to be potentially captured.

Findings. In total, there were 214 primary studies in our review that met the inclusion criteria. There were 2,338 variables described in the methods section of these articles. Of those variables, 70 were considered to be types of behavioral measures. That is, approximately three percent of variables studied were behavioral in nature. Interestingly, 173 studies (81%) had no behavioral measure, 26 studies had one behavioral measure, seven studies had two behavioral measures, and eight studies had between three and five behavioral measures. Forty four percent of studies that employed an experimental or quasi-experimental design included a measure of behaviors. For *Leadership Quarterly* specifically, 30 studies from articles published in 2019 met the inclusion criteria. Seven studies measured behaviors (three studies had one behavioral measure, two studies had two behavioral measures, and two studies had five behavioral measures each), whereas 23 did not. Across the total variables studied, a total of 4.48% of variables in *Leadership Quarterly* were behavioral in nature. In sum, the results from the systematic review suggest that a very small proportion of the research in leadership and the organizational behavior field study actual behaviors.

Of those studies that measured behaviors, the content of communication was most frequently used (24% of all behaviors coded). For example, Watts et al. (2019) looked at the quality, originality, and elegance of visionary plans that participants drafted up. Further, 17% of the behaviors coded were frequency counts, such as the number of sexually explicit jokes made (Mitchell et al., 2004) or the number of times women and men interrupted other students (Brooks, 1982). Other examples of behaviors coded include action/inaction decisions (6%; turnover [e.g., Russell & van Sell, 2012] or whether to engage in cooperative behavior [Kerr & Kaufman-Gilliland, 1994]), duration (e.g., amount of time to make a decision, length of speech [Brooks 1982; Frieder et al., 2016]). Most of these behaviors measured some aspect of leadership behavior (28% of behaviors coded), though topics such as performance (18%), voice behavior (7%), and work-family conflict (6%), among others, were also included. As some final examples from the area of leadership, Maran et al. (2019) leveraged eye tracking software and Obenauer and Langer (2019) studied leadership outcomes in the National Basketball Association. A more detailed list of the 70 behaviors we coded can be found on the second sheet of the online appendix.

Advancing behavior in leadership theories

To recap, in the paper thus far, we have illustrated that the study of behaviors is lacking in leadership and general organizational behavior research, which is problematic. Re-introducing the study of behavior into leadership research begins with addressing a variety of theoretical issues in the literature. This is because the absence of human behavior in leadership research is a theoretical problem first and foremost in terms of how we are conceptualizing concepts in theories. That is, we need to begin by formally separating perceptions, evaluations, and behaviors in leadership theories to properly guide later methodological decisions (see Figs. 1 and 2 for a demonstration). In the section that follows, we offer five theoretical recommendations for future research on leadership; we highlight these recommendations in Table 3.

Theoretical recommendation #1: Broaden conceptualizations of leader and follower behaviors

Our first recommendation is that leadership scholars should broaden their conceptualization of behaviors in order to fully develop theoretical advancements. To this end, we present in Table 4 a new framework of behaviors. In the table, we also overlay the results from the systematic search above to illustrate a sampling of what is currently being done in the literature. Though, not all of the behaviors

[2] Meta-analyses that did not include a reference section in the article or in an online appendix indicating the primary studies included in the review were excluded.

Table 3
Five theoretical recommendations for future research on leader (follower) behavior.

Recommendation	Description
1. Broaden the conceptualization of behavior	■ Consider a wider range of conceptualizations of leader (follower) behaviors in order to facilitate theoretical advancements (see Table 4 for a new theoretical framework of behaviors)
2. Avoid theoretical conflation	■ Explicitly define and specify concepts of perceptions, evaluations, behavior, and other related concepts (e.g., intentions, values)
3. Re-evaluate existing theories	■ Re-evaluate existing theories to ensure the concepts that they are composed of are properly specified
4. Conceptualize contamination/deficiency in concepts	■ Develop theory that accounts for both potential contamination and deficiency. Contamination and deficiency in the measurement of perceptions and evaluations are not always noise but could be theoretically relevant
5. Theorize inaction	■ Theorize the effects of both action and inaction in behaviors to fully understand the nature of leader (follower) behavior

in the proposed framework were represented in the systematic search. From a theoretical perspective, this allows for a broader, richer explanation of relevant phenomena. The behaviors we propose in the framework are various suggestions to think about leader and follower behaviors in different ways. In other words, we offer ways to incorporate behaviors that might not be obvious or apparent.

Behavior across timescales. The first theoretical approach covers the incorporation of time scales (Zaheer et al., 1999), which allows for both the consideration of behaviors and time. For instance, researchers can define and conceptualize behaviors by studying *rate,* which can be described as the frequency within an amount of time that something occurs (Aguinis & Bakker, 2021). As another example of a behavior type within the category of time, researchers might examine the exact *time* in which a behavior occurs. For instance, behavior can be characterized by the time of day at which a person wakes up or what time(s) a person eats meals throughout the day. Consider how, within signaling theory (Connelly et al., 2011), one might investigate the time at which a leader leaves work each day (a signal of the importance of work-life balance from a leader to followers). Also, within this theoretical approach, researchers might look at the passage of time, or *duration* (Shipp & Cole, 2015). We describe this approach as how long it takes an individual or team to complete a task (e.g., minutes, hours, and seconds). Six percent of the studies measuring behaviors in our systematic search used the passage of time to measure behavior. As an example, Maran et al. (2019) measured the length of time participants eyes were fixated on a video recording. Other recent empirical evidence indicates that the speed at which a leader makes and communicates a decision is related to evaluations of honesty (Van de Calseyde, Evans, & Demerouti, 2021).

Magnitude of behaviors. We give two examples under this category. The first is *frequency counts*. This involves the number of times a leader or follower engages in a behavior that is observed and objectively accounted for. For example, an algorithm might be used to objectively score the frequency of behaviors (see Pieterse et al., 2019 for a team-level performance dependent variable). Frequency counts were used often in the studies that measured behaviors included in our systematic search (17%). An example of the study of this behavior might be within role theory (Katz & Kahn, 1978), where the number of times an employee speaks up in a meeting might be used to examine leader emergence (MacLaren et al., 2021). Another

Table 4
A conceptual framework of leader (follower) behaviors.

Behavior type	Description	Examples
Theoretical approach #1: Leader (follower) behavior across timescales		
Rate (0%)	Frequency within an amount of time	Number of times a leader mentions various followers per minute during a conference call
Time (0%)	Time of day when behavior occurs	The time at which a leader arrives, and leaves work each day
Passage of time/ Duration (6%)	How long it takes to complete a task (e.g., minutes, hours, seconds) or speaking time	How long a follower works on a task
Theoretical approach #2: Magnitude of leader (follower) behaviors		
Frequency counts (17%)	Number of times an individual engaged in a behavior	Number of times an employee speaks up in a meeting (to predict leader emergence)
Intensity/ Magnitude (3%)	Strength of a behavior	Audio volume at which specific words are spoken in a number of interactions (e.g., offensive words in the case of abusive supervision)
Theoretical approach #3: Form of leader (follower) behavior		
Content of communication (24%)	Content of communications are objectively coded by humans or algorithms	Content coded from CEO letters to shareholders
Quality (1%)	How well a product is constructed, free from errors or deficiencies	The creativity of code created by software engineers
Physical form (0%)	What a behavior looks like physically	The shape or pattern of a leader's facial expressions
Accuracy (3%)	The extent to which an employee achieves a specific goal or target	Accuracy of decisions classifying items into categories
Action/inaction decisions (6%)	The decision to engage in a behavior or without engagement in the behavior	Decision to engage in extra role behavior (e.g., working past 5 pm when the workday has officially ended); Decision to remain with an organization or turnover
Theoretical approach #4: Behaviors in complex environments (from the complexity science literature)		
Stable behaviors (0%)	Behaviors are at a stable, single-point equilibria	A follower abstains from increasing his or her work effort
Periodic orbit behaviors (0%)	A regular sequence of states	The cycle of behaviors of followers that derives from an exogenous sequence of events
Chaotic behaviors (0%)	Behaviors are chaotic; they are extremely sensitive to initial conditions	A leader facilitates a political debate with followers on an internal company message board and it results in critiques from external stakeholders, such as investors
Complex behaviors (0%)	Initial patterns of behavior develop structures that begin to interact, and the patterns continue to evolve	A global pandemic results in an economic recession and workforces become remote opening up the door for the reconfiguration of how leaders and followers coordinate the completion of work

Note. Numbers in parentheses indicate the percentage of studies that included the respective behavior type, of those studies that included behavioral measures. These numbers do not sum to 100% as "other" behaviors are not included.

example of a behavior type that falls under this category is *intensity*, which we found in three percent of the studies that measured behaviors in our systematic search. This characterizes the strength of behavior. For instance, consider abusive supervision theory and the volume at which specific offensive words are spoken by a leader (Tepper, 2000) as a complementary approach to measuring follower evaluations of a leader (i.e., do followers evaluate these words to be offensive?).

Form of behavior. The third theoretical approach we wish to introduce describes the form of behaviors that scholars might build and test theory around. There are several examples of behavior types that might fall under this category. First, scholars might create and test theory around the behavior of communicating by examining the *content* of communication, which was the most frequently (24%) used behavior type, of the studies including behaviors from our systematic search. The content of the communication might be objectively coded by humans or algorithms. For instance, under agency theory (Dalton et al., 2007) content can be analyzed from executives' letters to shareholders (Short et al., 2010). Further, Weiss and Morrison (2019) coded decisions about the launch of a new product made by teams to determine the amount of innovation. As another behavior type, scholars might look at *quality,* which is characterized as how well a product is constructed free from errors or deficiencies. For example, theory might be tested to examine the role of prosocial motivation communicated by a leader in increasing the quality and creativity of code created by software engineers to serve a public good (Grant, 2008). Of the studies that measured behaviors in our search, only one percent used quality.

As a third type of the form of behavior, scholars might look at the *physical form* of a behavior. An example of this might be the study of the shape or pattern of a leader's facial expression in affective events theory (Weiss & Cropanzano, 1996). Another behavior type in this category is *accuracy*. This can be conceptualized as the extent to which followers achieve a specific goal or target. We found this type of behavior in three percent of the studies that measured behaviors in our search. For example, Lepine & Van Dyne (2001) measured the accuracy of a decision classifying aircrafts on a scale of nonthreatening to threatening. A final category of behavior may include *action or inaction decisions*. Six percent of the studies that measured behaviors in our systematic search included action or inaction decisions. Examples include the dichotomous decision (yes or no) to engage in an extra role behavior, such as when followers work beyond a contracted amount of time (e.g., Ernst et al., in press). As a second example, this might include the decision to leave an organization or to remain with an organization longer (e.g., turnover; Rubenstein et al., 2019; Russell & van Sell, 2012).

Behaviors in complex environments. Our fourth and final theoretical approach to broadening the conceptualization of behavior focuses on the study of leader and follower behavior in complex systems (Rosenhead et al., 2019). An environment is complex when "it consists of interdependent, diverse entities, and we assume that those entities adapt—that they respond to their local and global environments" (Page, 2009; p. 3). In other words, systems are typically characterized as complex when there are connections among people, diversity, interdependence exists, and people adapted. These features likely characterizes a large number of leader-follower interactions at all levels of organizations. Critically, complexity is an emergence phenomenon which is a concept likely familiar to leadership scholars. That is, macro-level phenomena materialize from interactions among lower-level phenomena that are nested (at the within-person-, dyadic, and/or unit-levels).

We have only begun to see leadership scholars touch upon complexity through the study of behavior. For example, behaviors are nested in teams and behaviors are nested within individuals (i.e., the intra-individual level of analysis). Thus, there is a need to consider process models in organizational behavior theories (Fischer et al., 2017). Drawing upon complexity science, we highlight examples of behavior in complex environments. First, there is the notion of *constant, or stable behaviors* (Page, 2009). Here, behaviors are at stable, single-point equilibria. An example of this is within transformational leadership theory where a follower does not identify with a leader and elects to abstain (an inaction) from increasing their work effort (Banks et al., 2018a). Second, there are *periodic orbit behaviors* which involve a regular sequence of states (Page, 2009). An example of this may occur where there is a cycle of behaviors among followers that derives from an exogenous sequence of events, such as an annual competition between departments for reaching sales goals (e.g., leaders seek to inspire followers to perform at high levels).

Third, there are *chaotic behaviors*. Here, behaviors are extremely sensitive to initial conditions. This example applies to chaos theory. An example is a leader facilitates a political debate with followers on an internal company message board and it results in critiques from external stakeholders, such as investors (Nicas, 2019). Fourth, there are *complex behaviors* (Page, 2009). Here, an initial pattern of behavior develops structures that begin to interact, and the patterns continue to evolve. For instance, a global pandemic results in an economic recession and workforces become remote, opening up the door for the reconfiguration of how leaders and followers coordinate the completion of work.

We wish to add the caveat that the behaviors that fall within complex environments may often be best studied with modeling. For instance, Samuelson et al. (2019) conducted two agent-based simulation studies to examine the influence of bias in hiring and developmental opportunities. This work built on previous modeling from Martell et al. (1996) which demonstrated how small amounts of bias in performance evaluations can contribute to the severe lack of women at upper echelons of organizations. Modeling approaches allow for strong control of all elements (diversity, interdependence, connections, adaptability) to understand behavior in emerging complex environments.

Overall, this behavioral framework represents some approaches to conceptualizing behaviors in leadership theories. Using this framework as a guide in future research can help to broaden how we build and test theory regarding leader and follower behaviors.

Theoretical recommendation #2: Avoid theoretical conflation

A second recommendation is to avoid conflation of concepts (see the description of Figs. 1 and 2 for a full discussion). This includes not treating concepts and measures as the same thing (Arthur & Villado, 2008). As an example, ethical leadership research has developed and tested theory in which behaviors are conceptualized to be distinct from follower evaluations (Banks et al., 2021a; 2021b). Scholars should explicitly develop and test theory which includes perceptions, evaluations, behavior, and other related concepts, such as intentions; though, they should also be purposeful in distinguishing them theoretically. Greater theoretical precision is needed in order to advance the domain of leadership and other areas in the organizational sciences. Addressing conflation may serve to help address other concerns in the area of leadership, such as concept redundancy (Banks et al., 2018a), or the idea that some leadership styles do not capture anything other than "do I like my leader" (Yammarino et al., 2020).

Theoretical recommendation #3: Re-evaluate existing theories

Our next recommendation builds upon these first two recommendations. That is, we call for existing leadership theories (and the concepts they are composed of) to be re-evaluated in order to address the issue of conflation raised throughout this paper. As one example, charismatic leadership, a very popular concept, has been plagued historically by conflation with other concepts, such as transformational leadership (Van Knippenberg & Sitkin, 2013). Moreover, charismatic leadership was primarily studied using self-report measures, which conflated evaluations of leaders with leader behaviors. The research also often suffered from endogeneity bias (Banks et al., 2017).

Critically, Antonakis et al. (2016) worked to redefine charismatic leadership based on signaling theory providing a much stronger theoretical foundation for the concept. This work "unconflated" transformational leadership from charismatic leadership and further separated evaluations of charismatic leadership from charismatic leadership behaviors. Building upon this work, a series of studies have been completed looking at actual charismatic leadership behavior as well as objective behavioral outcomes (e.g., Antonakis et al., 2011; Antonakis, d'Adda, Weber, & Zehnder, in press; Ernst et al., in press; Jacquart & Antonakis, 2015; Meslec et al. 2020). The literature on charismatic leadership is now based on improved theoretical grounding and has greater implications for leadership training and development as a result. That is, we now have a better understanding of how charisma, once thought to be a mystical quality, can be taught (Antonakis et al., 2012). We call for existing theories and concepts to be re-evaluated in a similar fashion to the work that was completed on charismatic leadership.

Theoretical recommendation #4: Conceptualize contamination/deficiency in concepts

In Figs. 1 and 2, we depict the potential for contamination and deficiency in leadership concepts. We suggest that these factors are not always noise but could be theoretically relevant. For instance, while familial status, home demands, social comparisons, and psychological contracts influence evaluations of leader behavior, this is theoretically relevant to know and conceptualize. As another example, scholars may approach inter-rater reliability of supervisor evaluations from a measurement perspective (Rothstein, 1990; Viswesvaran et al., 1996). We encourage leadership scholars to advance this perspective and develop theory around potential biases that can contribute to contamination or deficiency in perceptions, evaluations, and behaviors (see the description of Figs. 1 and 2 for examples). This helps leadership theories to be more precise and accurate. Developing theory around what might be perceived as "error" could help to create new theoretical insights.

As one illustrative example of contamination and deficiency, consider a CEO speech to shareholders at an annual meeting. This speech is likely composed by the CEO in conjunction with a speech writer. Hence, there are multiple voices involved (potential contamination), although the CEO delivers the speech and is the sole name associated with this work. We would say that regardless of who wrote the speech, the leader (e.g., CEO) is the one who is engaging in the signaling behavior. Hence, it is the behavior of the leader. However, leadership is a social influence process, and followers (stakeholders in the context of CEOs) are evaluating the signals. For instance, a follower might interpret a signal to be low cost (and less informative) unless the leader is taking a strong stance on a divisive issue. Observability of the signal depends on its strength, intensity, clarity, and visibility. Followers will also evaluate the reliability (credibility) of the signal. Signals sent by CEOs through very formal speeches in which a speech writer was involved (contamination) might be evaluated very differently than signals sent through spontaneous interactions, where the content of communication may be of more theoretical interest. Speeches may be deficient in a way if they do not capture these other signals. This is why it is so important to disentangle the behavior of leaders from the evaluations of followers and to be precise about what behavior is being studied (see Figs. 1 and 2). Many forms of seemingly similar behaviors are theoretically and practically relevant to study but may have different implications.

Theoretical recommendation #5: Theorize inaction

Finally, to fully understand the nature of leader and follower behavior, we should consider behavior as both action *and* inaction, in line with the definition of behavior that Levitis et al. (2009) offered. For instance, within transformational theory one might seek to understand when a leader fails to provide direction or exert social influence, as is the case with laissez-faire leader behavior. Or, within ethical leadership theory we might desire to understand when a leader chooses to engage in signaling behavior with some followers, but not others. As a final example, we might examine why a follower might decide to not help a coworker complete a task. Inaction is often ignored, but is theoretically relevant to consider (Levitis et al., 2009). That is, understanding why inaction occurs may be just as important as why actions do occur. This is especially relevant during the COVID-19 pandemic, where leaders across public and private organizations had important choices about what actions they should or should not take.

Methods-based recommendations for the study of behavior

In addition to the recommendations to build good leadership theory that incorporates and conceptualizes behavior, in this final section we highlight five key recommendations needed to methodologically improve the study of behavior in leadership research. It is likely that most leadership scholars have been trained and possess the skillset to implement these recommendations. However, due to time and publication pressure, such methodological techniques are not as commonly used (Antonakis, 2017; Antonakis et al., 2016; Day, 2014). We summarize these recommendations in Table 5.

Methodological recommendation #1: Align methods with theory

Methods, and in particular, the measures used in a study, should align with the theory to be examined. Bacharach (1989) made an important distinction between theory and data. Mainly, theory is NOT data. Rather, theory denotes a statement of the expected relationship between concepts, which are not observable, that are used to derive hypotheses, or statements of relationships among observable variables (Bacharach, 1989). Thus, variables (i.e., operationalizations of concepts) should match the concepts they are theoretically intended to measure. When theory and methods are misaligned, as is the case when researchers use a perceptual assessment of behaviors to represent behaviors, the conclusions drawn from the study are weak and prone to inference errors (MacKenzie et al., 2005). This is because the inferences one makes from the statistical conclusions of the data do not provide evidence for the concepts and theory in question. In sum, it is important to align methods with theory in order to promote a strong and accurate accumulation of knowledge.

Because much of the leadership literature conflates leader/follower behaviors with subjective evaluations, there is a need to re-evaluate the existing literature. This can be accomplished in two ways. First, scholars can work to address the conflation conceptually. Existing theories can be used in this regard. For example, signaling theory has been used in research on charismatic leadership and ethical leadership. This theory is useful because it separates leader behavior (signals) from follower evaluations and subsequent follower behavior (reactions to the signals and/or signals back to leaders). There are a number of articles which have recently worked to reconceptualize existing concepts using signaling theory in ways that separate behaviors from evaluations (Antonakis et al., 2016; Banks et al., 2021a).

Second, scholars can work to empirically identify leader behaviors. We recommend that a constant comparative approach be leveraged using data science techniques or traditional qualitative methods. This approach looks at emerging themes from new research while simultaneously comparing new themes to the extant literature. Traditional techniques such as Qualitative Content Analysis (QCA) or new approaches, like topic modeling (Banks et al., 2018b; Blei et al., 2003; Schmiedel et al., 2019), can be used to identify behaviors from videos, speeches, meeting transcripts, etc. Taxonomies of behaviors

Table 5
Five methodological recommendations for future research on leader and follower behaviors.

Recommendation	Description
1. Align methods with theory	Variables (i.e., operationalizations of concepts) should match the concepts they theoretically intend to measure. For instance, variables intended to measure leader (follower) behaviors should measure behaviors, rather than evaluations of behavior. This ensures inferences made from the empirical findings can inform the theory in question.
2. Consider reliability and validity	Reliability and validity of a measure can change the concept that one sets out to measure. This is particularly relevant for self-report and other ratings as well as descriptions of leader (follower) behavior. It is important to consider how a measure's reliability and validity vulnerabilities affect the focal concept.
3. Increase the use of objective measures	Objective measures of leader (follower) behavior are less prone to the reliability and validity concerns. As a result, they provide more "pure" evidence of behaviors and allow for a closer alignment of theory and methods. Artificial intelligence provides one means to accomplish this.
4. Triangulate using multiple measures when appropriate	Triangulation involves the process of using multiple measures or methods of measurement to study the same phenomenon. It can help to overcome the shortcomings and flaws of measures. If there is consensus or agreement among the findings from different methods or measures of the same phenomenon, there is increased confidence that the finding actually exists.
5. Use properly specified causal models	Experiments are the gold standard design approach for establishing support for a causal inference. When such a design is not possible, consider alternative approaches such as the use of an instrumental variable to allow for two-stage least squares regression analysis or difference-in-difference or a regression discontinuity approaches.

can then be created which account for existing literature while also accounting for concerns over conflation. These behaviors can then be investigated in causal models that explore the extent to which they are associated with a social influence process.

Methodological recommendation #2: Consider reliability and validity (contamination and deficiency)

As we demonstrated in Figs. 1 and 2, there are a number of potential concerns about the reliability and validity of the measures many leadership scholars use to assess behaviors (and evaluations). This is, in part, because unreliability and low validity can actually deem concepts measured to be different than the concepts they are intended to measure (Cortina et al., 2020; Heggestad et al., 2019). Validity concerns of contamination and deficiency can be present with self-report measures of behavior (e.g., memory recall and social desirability; Chan, 2009), other ratings of behaviors (Berry et al., 2012; Oh et al., 2011), as well as objective measures of behavior. With regard to reliability concerns, factors such as the time of day, the weather, or simply the wording of the scale items may change how people respond to items or how they interpret the environment (Heggestad et al., 2019). Some measures, such as questionnaires, are more likely to be influenced by transient factors, such as mood, time of day, and caffeine consumption (Le et al., 2009), leading to systematic errors that influence reliability. Thus, poor reliability and validity can quickly change the intended concept into an unrecognizable variable. As another example, questionnaire-based methods and behavioral measures can also suffer contamination, where the measurement of behavior also includes the measurement of another concept. Or, because of the complexity of leader (follower) behavior, a measure may miss an important element of the concept. In addition, from a statistical standpoint, reliability and validity concerns can also affect the magnitude of observed relationships, further influencing the conclusions that can be drawn from a set of results (Cortina et al., 2020). Failure to consider reliability and validity will not only lead to inaccurate relationships, but it may also lead to false conclusions, as the nature of the concept might actually change due to the unreliability and invalidity of the measures.

Regardless of one's approach to measure objective behaviors, perceptions, evaluations, intentions, values, etc., there are a number of factors that must be addressed in order to address deficiency and contamination. First, reporting accuracy can be influenced by the extent to which the items are objective and social desirability is minimized (Shaffer et al., 1986). Second, marker variables can be used to help understand differences in responses across followers that may contribute to an understanding of potential contamination and deficiency (Owens & Schoenfeldt, 1979). Third, scholars must consider attention when writing items. In a set of six studies, Hansbrough et al. (2021) demonstrated that item writing can play a role in whether episodic or semantic memories are activated. Episodic memories were associated with better estimation of targeted concepts as well as a reduction in other sources of bias. Approaches such as these may be associated with improved behavioral recall. Fourth, temporal framing is necessary to ensure that those who complete the questionnaire (e.g., leaders or followers) understand the context in which you are expecting descriptions of behavior over time (Stokes, Mumford, & Owens, 1989). In sum, it is important to consider the factors that influence reliability and validity of any measure.

Methodological recommendation #3: Increase the use of objective measures

Whenever possible, researchers should use objective measures to assess behaviors. While certainly not perfect, objective measures, such as frequency counts, accuracy, or the passage of time, are less prone to the reliability and some validity concerns mentioned above. We recognize that identifying behavioral data is harder than getting data through questionnaire measures. This is why this latter measure is the dominant approach. However, this certainly does not justify conflating or confusing behaviors with evaluations. So, we suggest that the first step to implement this recommendation is to stop theoretically conflating leader (follower) behaviors with evaluations of behaviors in the Introduction sections of articles. Experimental settings (e.g., a lab context) are a natural second step perhaps. Here, leader behavior can be manipulated and follower reactions (e.g., evaluations; behaviors) can be measured. Game-theoretic designs, common in the area of behavioral economics, that uncover actual choices and actions in context can also be leveraged (Zehnder et al., 2017). Podsakoff and Podsakoff (2019) provided a number of suggestions for how to overcome concerns regarding generalizability. One option, for instance, is to hire participants as temporary workers and pay them real wages and ask them to produce a real outcome (e.g., Antonakis et al., in press; Ernst et al., in press).

As a third step, there are a number of natural data sources of leader behavior, such as speeches, videos of leaders, and archival data from publicly traded companies, governments, and other public organizations. These datasets can be analyzed through traditional means (e.g., Ordinary Least Square regression), but can also be studied through more recent advancements. These include regression discontinuity designs (Stoker et al, 2019) and various data science techniques (Doornenbal et al., in press). Machine learning (Spisak et al., 2019) and artificial intelligence (LeCun et al., 2015) provide useful means to investigate behavior. It is important to ensure that proper steps are taken to ensure that concerns over endogeneity bias are reduced.

Still coding behavior directly does eliminate most of the inherent biases that are present in perceptual ratings, given that direct/objective ratings are neither convoluted nor confounded evaluative judgments (Banks, Fischer, Gooty, & Stock, 2020; Fischer, Tian, Lee, & Hughes, 2021). Hence it is easier to correct for endogeneity issues in behavioral rating via instrumental variable estimation, fixed-effects analysis, or using some other natural-experiment type design (Sieweke & Santoni, 2020; Siewke & Santoni, 2021). It is straightforward too to manipulate such behaviors (yet it impossible to directly manipulate evaluative judgments--one can only manipulate their causes, (Sajons, 2020).

Finally, there are more sensitive datasets that might be obtained from organizations. In some ways, gaining access is not a new challenge. Convincing an organization to distribute a (Fischer, Tian, Lee, & Hughes, 2021) survey and obtaining a sufficient response rate has never been an easy task. Here, the request of firms shifts, for instance, from distributing a survey to followers to asking firms to record Zoom meetings for later analysis using data science or traditional techniques. Again, we do not mean to imply this last avenue is easy, but it can be done. Moreover, scholars can continue to conduct traditional survey research, assuming their theory and methodological approach are aligned.

Many leadership scholars have been sufficiently trained to implement these steps. That is, they have been taught to study behaviors via traditional methods, such as experiments or to at least leverage questionnaires that are better aligned with theoretical questions. Where training may currently be lacking is in the area of data science. Machine learning algorithms and neural networks hold great promise for analyzing text, audio, and visual data in the area of leadership (Doornenbal et al., in press; Lee et al., in press). *Organizational Research Methods* has featured a number of special issues related to this. For instance, there is a special issue on video-based methods (https://journals.sagepub.com/doi/https://doi.org/10.1177/1094428117745649) . However, based on our systematic review of the literature, scholars have not begun to take advantage of these resources. As one recommendation to accelerate the use of data science, we encourage more collaboration with those in other disciplines, such as computer science. In fact, the National Science Foundation has existing programs that scholars can submit to for grant funding that can facilitate more behavioral based research that leverages both the organizational sciences (e.g., leadership) and data science (https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505620).

Methodological recommendation #4: Triangulate using multiple measures when appropriate

Triangulation involves the process of using multiple measures or methods of measurement to study the same phenomenon (Cox & Hassard, 2005; Jick, 1979). Triangulation can help overcome the shortcomings and flaws of measures, such as the reliability and validity concerns described above (Turner et al., 2017), particularly when objective measures are not available. If the findings regarding a particular phenomenon converge or agree across the different measures, it allows the researcher to draw stronger conclusions (Scandura & Williams, 2000). This is akin to evaluating convergent validity in the context of one's own study. Therefore, there is more confidence that the relationships observed are "real," rather than a methodological artifact (Jick, 1979).

Triangulation can take on many forms (Scandura & Williams, 2000). For instance, you could triangulate with different research methodologies, such as a laboratory study, a survey, and a simulation study (for an example of a five study approach see Ronay et al., 2019). Or one could use machine learning via neural networks to analyze leaders' speeches and videos and complement these efforts with a tightly controlled experimental design (LeCun et al., 2015) and interviews that levarage critical incidents. You could also triangulate with different sources of data, such as self-ratings versus other qualitative or quantitative ratings, ratings from multiple sources (such as in a 360-degree performance review), and/or the use of archival data, as examples (also consider a multitrait-multimethod approach). As one recent example, Hoogeboom et al. (in press) leveraged the use of skin conductance and video-based methods to examine leader behaviors. Data from these sources were then complemented with follower evaluations of leader effectiveness. Leader effectiveness was associated with higher levels of physiological arousal during both positive and negative relations-oriented behaviors. Overall, different methods of triangulation will be appropriate for different research questions. Regardless, triangulation helps researchers to find evidence closer to "truth" by overcoming the flaws inherent in various approaches (Cox & Hassard, 2005).

Methodological recommendation #5: Use properly specified causal models (i.e., mitigate endogeneity bias)

It is important to use properly specified causal models to mitigate concerns of endogeneity bias, which occurs when "the effect of x on y cannot be interpreted because it includes omitted causes" (Antonakis et al., 2010; p. 1087). This allows for a complete understanding of the causes and consequences of behaviors in organizational behavior theories (Hill, Johnson, Greco, O'Boyle, & Walter, 2021). Measurement error, as described above, and common method bias, of which self-reports are a common component, often lead to endogeneity problems (Clougherty et al., 2016). Other problems such as omitted variables and simultaneity can contribute to endogeneity bias, such as when follower behaviors cause leader behaviors in the study of leadership (Güntner et al., 2020).

Endogeneity issues are problematic because they bias effects; in such instances, the predictor variable is correlated with the error term in the model, which influences the true effect of the predictor on the outcome (Antonakis et al., 2010; Hill et al., 2021). Endogeneity bias can be easily mitigated or removed through the use of experimental designs where the independent variable is manipulated (Podsakoff & Podsakoff, 2019). Experiments are the gold standard design for allowing for causal inferences. However, other quasi-experimental techniques, such as difference-in-difference and regression discontinuity approaches can also be used to study behaviors when there are exogenous shocks (for examples see Stoker et al., 2019). For instance, difference-in-difference models could be leveraged in which a group experiences a type of exogeneous stimulus which is then compared to another group that did not receive such a treatment and this occurs over a period of time (Hill, Johnson, Greco, O'Boyle, & Walter, 2021). Endogeneity concerns can also be mitigated by collecting data on the independent and dependent variables from different sources or the use of an instrumental variable (for reviews see Antonakis, 2017; Antonakis et al., 2010). Instrumental variables can then be used in two-stage least squares (2SLS) regression analyses which provide for stronger causal inferences.

To be clear, the use of objective behavioral measures and/or collecting data from multiple sources alone does not inherently eliminate concerns about endogeneity bias. If the independent variable is not truly exogenous, it may correlate with omitted variables for instance. In this case, endogeneity bias is just as much a concern as in the typical questionnaire study. For instance, one could use deep neural networks to measure leader behavior, however, causal inferences may still be problematic (for a review and suggested remedies see Lee et al., in press). Thus, appropriate steps still need to be taken. Still coding behavior directly does eliminate most of the inherent biases that are present in perceptual ratings given that direct/objective ratings are neither convoluted nor confounded evaluative judgments (Banks, Fischer, Gooty, & Stock, 2020; Fischer, Tian, Lee, & Hughes, 2021). Hence it is easier to correct for endogeneity issues in behavioral rating via instrumental variable estimation, fixed-effects analysis, or using some other natural-experiment type design (Sieweke & Santoni, 2020). It is straightforward too to manipulate such behaviors (yet it impossible to directly manipulate evaluative judgments--one can only manipulate their causes, Sajons, 2020).

Are survey methods ever appropriate to study behavior?

While the current work has advocated for an increase in the use of behavioral measures, we have also mentioned on a number of occasions that there are caveats, limitations, or contingency factors in the use of behavioral measures. There is a lot of work on behavior in animal ecology (Berger-Tal et al., 2011; Creel & Creel, 1995) and behavioral economics literatures (Chapman, Milkman, Rand, Rogers, & Thaler, 2021; Thaler, 2018) that has done an excellent job investigating behavior. However, these same approaches often reflect a black-box in cognition, which provides an argument in favor of more perceptual measures. What the current work proposes is that scholars should attend equally to measures of behavior as well as psychological measures. Hence, we close by answering the question, are survey methods ever appropriate to study behavior (in addition to measuring psychological constructs)?

First and foremost, this is a theoretical point of consideration. That is, if the focal concept of interest is psychological, then a survey-based approach will absolutely be superior. Perception can be reality, resulting in multiple realities according to followers (Park & Sturman, in press). Hence survey methods are clearly ideal for studying perceptions, evaluations, and other psychological constructs. For instance, if a leader engages in a particular behavior, we may be explicitly interested in how followers evaluate that behavior. Scholars might examine gender bias and investigate the extent to which gender moderates the relation between a leader's behavior and the evaluations of followers (Braddy et al., 2020).

In addition to theoretical concerns, there may also be instances when a survey-based measure is more appropriate for methodological reasons. While survey-based methods may be prone to bias, as mentioned in Figs. 1 and 2, there is certainly potential for contamination or deficiency in behavioral-based measures. In fact, there may be instances in which followers' evaluation of a leader's behavior is more useful than a more direct measurement of the behavior which may suffer from severe deficiency. That is, the followers may have more information, and hence, their subjective evaluations better advance theory in a complex environment than a weak or limited behavioral measure. This may especially be the case when the desired behavior to be studied only occurs infrequently, such as with destructive leadership behaviors (a leader says something destructive in a private meeting). These behaviors may only rarely occur because one who engages in them all the time may not be a leader for long. Hence, direct evaluation by followers may be the only option at times to advance theory (e.g., Hill & Kintigh, 2009). As mentioned in Methodological Recommendation #4, the best designed studies may actually use a combination of behavioral and evaluation methods to triangulate on the phenomenon of interest. This way, evaluations and more objective measures can be complements to each other as a means to reduce bias either in the evaluations or the objective measures.

A call for a revolution in leadership research and beyond

Many readers may be familiar with the fable titled *The Emperor's New Clothes* in which an emperor is fooled into walking through the streets of his city without any clothes. His subjects are hesitant to say anything until finally a young child humbly states out loud what all the adults are thinking. That is, the emperor has no clothes. We think that this fable transfers to the current circumstance in the area of organizational behavior. Collectively as a field, we are aware of the fact that we are not studying actual behaviors, though we are reluctant to draw attention to the problem. Whether it is because of underrating the importance of behavior, methodological convenience (Antonakis et al., 2016; Day, 2014; Fischer et al., 2020), or a lack of awareness of potential bias, behaviors are hardly studied in leadership research or organizational behavior more broadly. Moreover, there is a conflation of behavior with perceptions and evaluations if our theories do not specify the differences.

There are three primary revolutions occurring in organizational sciences research right now: (1) a causal revolution, (2) a machine learning revolution, and (3) an open science revolution (Haveman et al., 2019). In the extant work, we have advocated for a fourth revolution in management: a behavior revolution. The study of behavior is of critical importance for the advancement of leadership theory, evidence-based practice, and policy making. However, we have reached a critical turning point in the evolution of this field. Despite its implications, a behavior revolution is not as utopian as it might seem. Rather, it returns us to the roots of management as a field, when social psychologists like Lewin et al. (1939), Bales (1950), and others studied true behaviors. Seeing as behavioral research was done then, it is certain to be feasible now.

Author note

We would like to thank Thomas Fischer, Janaki Gooty, Eric Heggestad, and Scott Tonidandel for suggestions related to this work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adler, A. B., Thomas, J. L., & Castro, C. A. (2005). Measuring up: Comparing self-reports with unit records for assessing soldier performance. *Military Psychology, 17*(1), 3–24.

Aguinis, H., & Bakker, R. M. (2021). Time is of the essence: Improving the conceptualization and measurement of time. *Human Resource Management Review, 31*, 100763.

Aguinis, H., Cummings, C., Ramani, R. S., & Cummings, T. G. (2019). An A is an A:" Design thinking and our desired future. *Academy of Management Perspectives, 33*, 264–266.

Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal, 40*, 1291–1315.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*(5), 888–918.

Andersson, L., Jackson, S. E., & Russell, S. V. (2013). Greening organizational behavior: An introduction to the special issue. *Journal of Organizational Behavior, 34*, 151–155.

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., Fenley, M., & Liechti, S. (2011). Can charisma be taught? Tests of two interventions. *Academy of Management Learning and Education, 10*, 374-396.

Antonakis, J., Fenley, M., & Liechti, S. (2012). Learning charisma. Transform yourself into the person others want to follow. *Harvard Business Review, 90,* 127-130, 147.

Antonakis, J., Bastardoz, N., Liu, Y., & Schriesheim, C. A. (2014). What makes articles highly cited? *The Leadership Quarterly, 25*(1), 152–179.

Antonakis, J., d'Adda, G., Weber, R., Zehnder, C. (in press). Just words? Just speeches? On the economic value of charismatic leadership. Management Science.

Antonakis, J., Bastardoz, N., Jacquart, P., & Shamir, B. (2016). Charisma: An ill-defined and ill-measured gift. *Annual Review of Organizational Psychology and Organizational Behavior, 3*(1), 293–319.

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly, 28*(1), 5–21.

Arthur, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*(2), 435–442.

Ashkanasy, N. M. (2013). Onward and upward: Reviewing the past, present, and future of JOB. *Journal of Organizational Behavior, 34*(1), 1–5.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review, 14*(4), 496. https://doi.org/10.2307/258555.

Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups.* Cambridge, Massachusetts.: Addison-Wesley.

Banks, G. C., Engemann, K. N., Williams, C. E., Gooty, J., McCauley, K. D., & Medaugh, M. R. (2017). A meta-analytic review and future research agenda of charismatic leadership. *The Leadership Quarterly, 28*(4), 508–529.

Banks, G.C., Fischer, T., Gooty, J., & Stock, G. (2021). Ethical leadership: Mapping the terrain for concept cleanup and a future research agenda. The Leadership Quarterly, 32. 101471.

Banks, G. C., Fischer, T., Gooty, J., & Stock, G. (2021a). Mapping the terrain of ethical leadership (behavior): Concept cleanup and future research agenda. *The Leadership Quarterly, 32* 101471.

Banks, G. C., Ross, R. L., Toth, A., Tonidandel, S., Goloujeh, A., & Wenwen, D. (2021b). A signaling theory approach to ethical leadership. Paper presented at the annual meeting of the Academy of Management, Philadelphia.

Banks, G. C., Gooty, J., Ross, R. L., Williams, C. E., & Harrington, N. T. (2018a). Construct redundancy in leader behaviors: A review and agenda for the future. *The Leadership Quarterly, 29*(1), 236–251.

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018b). A review of best practice recommendations for text-analysis in R (and a user friendly app). *Journal of Business and Psychology, 33*(4), 445–459.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*(4), 396–403.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organizational Research Methods, 6*, 440–464.

Becker, J. A., Ellevold, B., & Stamp, G. H. (2008). The creation of defensiveness in social interaction II: A model of defensive communication among romantic couples. *Communication Monographs, 75*, 86–110.

Berger-Tal, O., Polak, T., Oron, A., Lubin, Y., Kotler, B. P., & Saltz, D. (2011). Integrating animal behavior and conservation biology: A conceptual framework. *Behavioral Ecology, 22*(2), 236–239.

Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other-reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology, 97*(3), 613–636.

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology, 89*(5), 901–910.

Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science, 29*(7), 1084–1093.

Braddy, P. W., Sturm, R. E., Atwater, L., Taylor, S. N., & McKee, R. A. (2020). Gender bias still plagues the workplace: Looking at derailment risk and performance with self-other ratings. *Group & Organization Management, 45*(3), 315–350.

Brooks, V. R. (1982). Sex differences in student dominance behavior in female and male professors' classrooms. *Sex Roles, 8*, 683–690.

Carpenter, N. C., Berry, C. M., & Houston, L. (2014). A meta-analytic comparison of self-reported and other-reported organizational citizenship behavior. *Journal of Organizational Behavior, 35*(4), 547–574.

Chan, D. (2009). So why ask me? Are self-report data really that bad. Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences, 309-336.

Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology, 90*(5), 928–944.

Chapman, G., Milkman, K. L., Rand, D., Rogers, T., & Thaler, R. H. (2021). Nudges and choice architecture in organizations: New frontiers. *Organizational Behavior and Human Decision Processes, 163*, 1–3.

Clougherty, J. A., Duso, T., & Muck, J. (2016). Correcting for self-selection based endogeneity in management research: Review, recommendations and simulations. *Organizational Research Methods, 19*(2), 286–347.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management, 37*(1), 39–67.

Cortina, J. M., Sheng, Z., Keener, S. K., Keeler, K. R., Grubb, L. K., Schmitt, N., Tonidandel, S., Summerville, K. M., Heggestad, E. D., & Banks, G. C. (2020). From alpha to omega and beyond!: A look at the past, present, and (possible) future of psychometric soundness in the Journal of Applied Psychology. *Journal of Applied Psychology, 105*(12), 1351–1381.

Cox, J. W., & Hassard, J. (2005). Triangulation in organizational research: A re-presentation. *Organization, 12*(1), 109–133.

Creel, S., & Creel, N. M. (1995). Communal hunting and pack size in African wild dogs, *Lycaon* picture. *Animal Behaviour, 50*, 1325–1339.

Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology, 90*(6), 1241–1255.

Dalton, D. R., Hitt, M. A., Certo, S. T., & Dalton, C. M. (2007). The fundamental agency problem and itsmitigation: Independence, equity, and the market for corporate control. *Academy of Management Annals, 1*, 1–64.

Day, D. V. 2014. The future of leadership: Challenges and prospects.

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology, 17*, 245–260.

Doornenbal, B. M., & Spisak, B. R., van der Laken, P. A. (in press). Opening the black box: Uncovering the leader trait paradigm through machine learning. *The Leadership Quarterly.*

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

Ernst, B., Banks, G. C., Loignon, A. C., Frear, K. A., Williams, C. E., Arciniega, L. M., Gupta, R. K., Kodydek, G., Subramanian, D. (in press). Investigating charismatic leadership and signaling theory: A prospective meta-analysis in five countries. *The Leadership Quarterly.*

Fischer, T., Dietz, J., & Antonakis, J. (2017). Leadership process models: A review and synthesis. *Journal of Management, 43*(6), 1726–1753.

Fischer, T., Hambrick, D. C., Sajons, G. B., & Van Quaquebeke, N. (2020). Beyond the ritualized use of questionnaires: Toward a science of actual behaviors and psychological states. *The Leadership Quarterly.*.

Fischer, T., Tian, A. W., Lee, A., & Hughes,, D. J. (2021). Abusive Supervision: A Systematic Review and Fundamental Rethink. *The Leadership Quarterly, 31*, 101348.

Frieder, R. E., Van Iddekinge, C. H., & Raymark, P. H. (2016). How quickly do interviewers reach decisions? An examination of interviewers' decision-making time across applicants. *Journal of Occupational and Organizational Psychology, 89*(2), 223–248.

Furnham, A., & Stringfield, P. (1998). Congruence in job-performance ratings: A study of 360 feedback examining self, manager, peers, and consultant ratings. *Human Relations, 51*(4), 517–530.

Güntner, A. V., Klonek, F. E., Lehmann-Willenbrock, N., & Kauffeld, S. (2020). Follower behavior renders leader behavior endogenous: The simultaneity problem, estimation challenges, and solutions. *The Leadership Quarterly, 31*(6), 101441. https://doi.org/10.1016/j.leaqua.2020.101441.

Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment, 11*(4), 340–344.

Grant, A. M. (2008). Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. *Journal of Applied Pschology, 93*(1), 48–58.

Hansbrough, T. K., Lord, R. G., Schyns, B., Foti, R. J., Liden, R. C., & Acton, B. P. (2021). Do you remember? Rater memory systems and leadership measurement. *The Leadership Quarterly, 32*(2), 101455. https://doi.org/10.1016/j.leaqua.2020.101455.

Hammer, L. B., Kossek, E. E., Yragui, N. L., Bodner, T. E., & Hanson, G. C. (2009). Development and validation of a multidimensional measure of family supportive supervisor behaviors (FSSB). *Journal of Management, 35*(4), 837–856.

Haveman, H. A., Mahoney, J. T., & Mannix, E. (2019). Editor's comments: The role of theory in management research. *Academy of Management Review, 44*(2), 241–243.

Heggestad, E. D., Scheaf, D. J., Banks, G. C., Monroe Hausfeld, M., Tonidandel, S., & Williams, E. B. (2019). Scale adaptation in organizational science research: A review and best-practice recommendations. *Journal of Management, 45*(6), 2596–2627.

Heneman, H. G., Heneman, R. L., & Judge, T. A. (1997). *Staffing organizations*. Madison, WI: Mendota House/Irwin.

Henriques, G., & Michalski, J. (2020). Defining Behavior and its Relationship to the Science of Psychology. *Integrative Psychological Behavioral Science, 54*(2), 328–353.

Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H., & Walter, S. L. (2021). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management, 47*, 104–143.

Hill, K., & Kintigh, K. (2009). Can anthropologists distinguish good and poor hunters? Implications for hunting, hypotheses, sharing conventions, and cultural transmission. *Current Anthropology, 50*(3), 369–378.

Hoogeboom, M. A., Saeed, A., Noordzij, M. L., Wilderom, C. P. (in press). Physiological arousal variability accompaning relations-orieted behaviors of effective leaders: Triangulating skin conductance, video-based beahvior coding and perceived effectiveness. *The Leadership Quarterly.*

Jacquart, P., & Antonakis, J. (2015). When does charisma matter for top-level leaders? Effect of attributional ambiguity. *Academy of Management Journal, 58*(4), 1051–1074.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*(4), 602. https://doi.org/10.2307/2392366.

Johns, G., & Miraglia, M. (2014). *The reliability, validity, and accuracy of self-reported absenteeism from work: a meta-analysis.* Paper presented at the Academy of Management Proceedings.

Johnson, B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. *Handbook of Mixed Methods in Social and Behavioral Research*, 297–319.

Judge, T. A., Cable, D. M., Colbert, A. E., & Rynes, S. L. (2007). What causes a management article to be cited—Article, author, or journal? *Academy of Management Journal, 50*(3), 491–506.

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations*. New York, NY: Wiley.

Kerr, N. L., & Kaufman-Gilliland, C. M. (1994). Communication, commitment, and cooperation in social dilemma. *Journal of Personality and Social Psychology, 66*(3), 513–529.

Le, H., Schmidt, F. L., & Putka, D. J. (2009). The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organizational Research Methods, 12*(1), 165–200.

LeBaron, C., Jarzabkowski, P., Pratt, M. G., & Fetzer, G. (2018). An introduction to video methods in organizational research. In: SAGE Publications Sage CA: Los Angeles, CA.

LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology, 92*(1), 1–16.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.

Lee, A., Inceoglu, I., Hauser, O., & Greene, M. (in press). Determining causal relationships in leadership research using Machine Learning: The powerful synergy of experiments and data science. *The Leadership Quarterly*, 101426.

LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology, 87*(1), 52–65.

LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with big five personality characteristics and cognitive ability. *Journal of Applied Psychology, 86*(2), 326–336.

Levitis, D. A., Lidicker, W. Z., & Freund, G. (2009). Behavioural biologists do not agree on what constitutes behaviour. *Animal Behaviour, 78*(1), 103–110.

Lewin, K., Lippitt, R., & White, R. K. (1939). Patterns of aggressive behavior in experimentally created "social climates". *The Journal of Social Psychology, 10*(2), 269–299.

Liao, H., Toya, K., Lepak, D. P., & Hong, Y. (2009). Do they see eye to eye? Management and employee perspectives of high-performance work systems and influence processes on service quality. *Journal of Applied Psychology, 94*(2), 371–391.

Liu, W., Tangirala, S., Lee, C., & Parker, S. K. (2019). New directions for exploring the consequences of proactive behaviors: Introduction to the special issue. *Journal of Organizational Behavior, 40*, 1–4.

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Pschology, 90*(4), 710–730.

MacLaren, N. G., Yammarino, F. J., Dionne, S. D., Sayama, H., Mumford, M. D., Connelly, S., & Kulkarni, A. (2021). Testing the babble hypothesis: Speaking time predicts leader emergence in small groups. *The Leadership Quarterly, 31*, 101409.

Maran, T., Furtner, M., Liegl, S., Kraus, S., & Sachse, P. (2019). In the eye of a leader: Eye-directed gazing shapes perceptions of leaders' charisma. *The Leadership Quarterly, 30*(6), 101337. https://doi.org/10.1016/j.leaqua.2019.101337.

Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-female differences: A computer simulation. *American Psychologist, 51*(2), 157–158.

Meslec, N., Curseu, P. L., Fodor, O. C., & Kenda, R. (2020). Effects of charismatic leadership and rewards on individual performance. *The Leadership Quarterly, 31*(6), 101423. https://doi.org/10.1016/j.leaqua.2020.101423.

Meyer, H. H. (1975). The pay-for-performance dilemma. *Organizational Dynamics, 3*(3), 39–50.

Mitchell, D., Hirschman, R., D. J., A., & Lilly, R. S. (2004). A laboratory analogue for the study of peer sexual harassment. *Psychology of Women Quarterly, 28*(3), 194–203.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*(4), 475–480.

Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychology Measurement, 11*(1), 1–31.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263–280.

Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection Assessment, 13*(4), 250–260.

Nicas, J. (2019). Google tries to corral its staff after ugly internal debates. *The New York Times.* Retrieved from https://www.nytimes.com/2019/08/23/technology/google-culture-rules.html.

Obenauer, W. G., & Langer, N. (2019). Inclusion is not a slam dunk: A study of differential leadership outcomes in the absence of a glass cliff. *The Leadership Quarterly, 30*(6), 101334. https://doi.org/10.1016/j.leaqua.2019.101334.

Oh, I.-S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology, 96*(4), 762–773.

Ostroff, C., & Bowen, D. E. (2016). Reflections on the 2014 decade award: Is there strength in the construct of HR system strength? *Academy of Management Review, 41*(2), 196–214.

Owens, W. A. (1976). Background data. In *Handbook of Industrial Organizational Psychology* (Vol. 3, pp. 61-138).

Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. *Journal of Applied Psychology, 64*(5), 569–607.

Page, S. E. (2009). *Understanding complexity: Course guidebook*. Chantilly, VA: The Teaching Company.

Park, S. & Sturman, M.C. (in press). When peception is reality, there is more than one reality: The formaion and effects of pay-for-performance perceptions. *Personnel Psychology*.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598–609.

Pieterse, A. N., Hollenbeck, J. R., van Knippenberg, D., Spitzmüller, M., Dimotakis, N., Karam, E. P., & Sleesman, D. J. (2019). Hierarchical leadership versus self-management in teams: Goal orientation diversity as moderator of their relative effectiveness. *The Leadership Quarterly, 30* 101343.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. *Organizational Research Methods, 19*(2), 159–203.

Podsakoff, P. M., & Podsakoff, N. P. (2019). Experimental designs in management and leadership research: Strengths, limitations, and recommendations for improving publishability. *The Leadership Quarterly, 30*(1), 11–33.

Robbins, S. P., & Judge, T. A. (2008). *Robbins, Stephen P. & Judge, Timothy A. Essentials of Organizational Behavior, 13th edition. Upper Saddle River, NJ: Prentice Hall* (13th ed.). Upper Saddle River, NJ: Prentice Hall.

Ronay, R., Oostrom, J. K., Lehmann-Willenbrock, N., Mayoral, S., & Rusch, H. (2019). Playing the trump card: Why we select overconfident leaders and why it matters. *The Leadership Quarterly., 30*(6), 101316. https://doi.org/10.1016/j.leaqua.2019.101316.

Rosenhead, J., Franco, L. A., Grint, K., & Friedland, B. (2019). Complexity theory and leadership practice: A review, a critique, and some recommendations. *The Leadership Quarterly, 30*(5), 101304. https://doi.org/10.1016/j.leaqua.2019.07.002.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology, 75*(3), 322–327.

Rubenstein, A. L., Kammeyer-Mueller, J. D., Wang, M. o., & Thundiyil, T. G. (2019). "Embedded" at hire? Predicting the voluntary and involuntary turnover of new employees. *Journal of Organizational Behavior, 40*(3), 342–359.

Russell, C. J., & Sell, M. V. (2012). A closer look at decisions to quit. *Organizational Behavior and Human Decision Processes, 117*(1), 125–137.

Sajons, G. B. (2020). Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables. *The Leadership Quarterly, 31*, 101338.

Samuelson, H. L., Levine, B. R., Barth, S. E., Wessel, J. L., & Grand, J. A. (2019). Exploring women's leadership labyrinth: Effects of hiring and developmental opportunities on gender stratfication. *The Leadership Quarterly, 30* 101314.

Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal, 43*, 1248–1264.

Schmiedel, T., Müller, O., & vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods, 22*(4), 941–968.

Shaffer, G. S., Saunders, V., & Owens, W. A. (1986). Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. *Personnel Psychology, 39*(4), 791–809.

Shipp, A. J., & Cole, M. S. (2015). Time in individual-level organizational studies: What is it, how is it used, and why isn't it exploited more often? *Annual Review. Organizational Psychology and Organizational Behavior, 2*(1), 237–260.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational Research Methods, 13*(2), 320–347.

Sieweke, J., & Santoni, S. (2020). Natural experiments in leadership research: An introduction, review, and guidelines. The Leadership Quarterly, 31(1), 101338.

Siewke, J., & Santoni, S. (2021). Natural experiments in leadership research: An introduction, review, and guidelines. *The Leadership Quarterly, 31*, 101338.

Spector, P. E. (1994). Using self-report questionnaires in OB research: A comment on the use of a controversial method. *Journal of Organizational Behavior, 15*(5), 385–392.

Spisak, B. R., van der Laken, P. A., & Doornenbal, B. M. (2019). Finding the right fuel for the analytical engine: Expanding the leader trait paradigm through machine learning? *The Leadership Quarterly, 30*(4), 417–426.

Steiner, D. D., & Rain, J. S. (1989). Immediate and delayed primacy and recency effects in performance evaluation. *Journal of Applied Pschology, 74*(1), 136–142.

Stewart, S. M., Bing, M. N., Davison, H. K., Woehr, D. J., & McIntyre, M. D. (2009). In the eyes of the beholder: A non-self-report measure of workplace deviance. *Journal of Applied Psychology, 94*(1), 207–215.

Stoker, J. I., Garretsen, H., & Soudis, D. (2019). Tightening the leash after a threat: A multi-level event study on leadership behavior following the financial crisis. *The Leadership Quarterly, 30*(2), 199–214.

Stokes, G. S., Mumford, M. D., & Owens, W. A. (1989). Life history prototypes in the study of human individuality. *Journal of Personality, 57*, 509–545.

Tepper, B. J. (2000). Consequences of abusive supervision. *Academy of Management journal, 43*, 178–190.

Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: path analyses based on meta-analytic findings. *Personnel Psychology, 46*, 259-293.

Thaler, R. H. (2018). From cashews to nudges: The evolution of behavioral economics. *American Economic Review, 108*(6), 1265–1287.

Turner, S. F., Cardinal, L. B., & Burton, R. M. (2017). Research design for mixed methods: A triangulation-based framework and roadmap. *Organizational Research Methods, 20*(2), 243–267.

Van de Calseyde Evans, A. M., & Demerouti, E. (2021). Leader decision speed as a signal of honesty. *The Leadership Quarterly, 31*, 101442.

van Knippenberg, D., & Sitkin, S. B. (2013). A critical assessment of charismatic—transformational leadership research: Back to the drawing board? *Academy of Management Annals, 7*(1), 1–60.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Pschology, 81*(5), 557–574.

Weiss, H. M., & Cropanzano, R. (1996). Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. In B. M. Staw & L. L. Cummings (Eds.). *Research in organizational behavior: An annual series of analytical essays and critical reviews* (Vol. 18, pp. 1–74). Elsevier Science.

Weiss, M., & Morrison, E. W. (2019). Speaking up and moving up: How voice can enhance employees' social status. *Journal of Organizational Behavior, 40*(1), 5–19.

Watts, L. L., Steele, L. M., & Mumford, M. D. (2019). Making sense of pragmatic and charismatic leadership stories: Effects on vision formation. *The Leadership Quarterly, 30*(2), 243–259.

Yammarino, F.J., Cheong, M., Kim, J., & Tsai, C.-Y. (2020). Is leadership more than "I like my boss"? *Research in Personnel & Human Resources Management, 38*, 1-55. ISSN: 0742-7301/doi:10.1108/S0742-730120200000038003

Zaheer, S., Albert, S., & Zaheer, A. (1999). Time scales and organizational theory. *Academy of Management Review, 24*(4), 725. https://doi.org/10.2307/259351.

Zehnder, C., Herz, H., & Bonardi, J.-P. (2017). A productive clash of cultures: Injecting economics into leadership research. *The Leadership Quarterly, 28*(1), 65–85.