The Leadership Quarterly 34 (2023) 101710

Contents lists available at ScienceDirect

The Leadership Quarterly

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

Eight puzzles of leadership science

George C. Banks *

UNC Charlotte, United States

ARTICLE INFO

Keywords:
Leadership
Quantitative
Qualitative
Micro
Macro

ABSTRACT

In the early 1900s, David Hilbert introduced a set of 23 mathematical problems. These problems caught the imagination of mathematicians around the world for the coming century and beyond. The advantage of such defined scientific puzzles is to galvanize coordinated efforts to pursue key scientific questions, whose answers can also address the grand challenges faced by society. Consequently, science can advance more quickly in a theory-driven, rigorous, open and collaborative manner, with meaningful implications for stakeholders. In the current editorial, I propose eight puzzles that could similarly motivate and guide leadership science. Whereas this list is not exhaustive--rather it is just eight starting points--beginning to solve such puzzles will accelerate our science and deliver on shared promises to our stakeholders. I then present actionable guidelines necessary to solve these puzzles and conclude by discussing the role *The Leadership Quarterly* will play in facilitating the pursuit of solutions to these puzzles.

Introduction

> *Who of us would not be glad to lift the veil behind which the future lies hidden; to cast a glance at the next advances of our science and at the secrets of its development during future centuries?* -Hilbert, 1902 (p. 437)

Leadership is a dynamic social influence process (Day & Antonakis, 2013), and it is an interdisciplinary academic field that is of great importance in all spheres of our lives from government and education to sports and religion (Day, Fleenor, Atwater, Sturm, & McKee, 2014). Scientists study leadership across disciplines such as in anthropology, biology, economics, management, political science, psychology, and sociology. In the study of leadership, scientists have discovered a tremendous amount of knowledge to explain the phenomenon. Nevertheless, much of what we know about leadership remains beyond our comprehension. Our job is to build on past accomplishments to reach new heights and achieve that which has not yet been previously achieved (Gardner et al., 2020).

In the early 20th century, Hilbert (1902), published 23 mathematical problems. These problems were presented as a set of goals and served to galvanize mathematicians in terms of intensity, direction, and persistence, while also promoting collaboration and a shared understanding (Fang & Casadevall, 2015; Locke & Latham, 2002; Robbins & Judge, 2017). In the current editorial, I present a set of scientific puzzles as goals to be solved, with the similar hope of energizing and focusing leadership science in the years to come. First, I review important societal issues and the role of scientific puzzles as a means to accomplish them. The scientific method is necessary here to advance the study of leadership in the same way science has helped us to accelerate other breakthroughs, such as vaccines for the COVID-19 pandemic (Wulff et al., 2023). Second, I introduce and describe eight puzzles, which can only be solved via rigorous basic and applied research. I also discuss actionable guidelines for institutions and individual scholars to solve these puzzles. Finally, in this position paper I conclude by explain the role *The Leadership Quarterly* (LQ) will play in facilitating the pursuit of solutions to these puzzles.

Grand challenges and scientific puzzles

Grand challenges are very large-scale goals or problems faced by society (Banks, Pollack, et al., 2016). Classic examples include establishing women's right to vote just over 100 years ago in the United States (Doepke, Tertilt, & Voena, 2012), the goal to place an astronaut on the moon (Stenger, 2001), the vision to end apartheid in South Africa (Mandela, 1993), and the global effort to map the human genome (Stephens, Cavanaugh, Gradie, Mador, & Kidd, 1990). The notion of grand challenges has entered into a number of specific domains such as management (George, Howard-Grenville, Joshi, & Tihanyi, 2016), computer science (Mertens & Barbian, 2015), epidemiology (Daar et al., 2007), mental (Collins et al., 2011) and global health (Varmus et al., 2003) as well as archeology (Kintigh et al., 2014) and energy sciences (Manley, Anastas, & Cue, 2008). The grand challenges of

Author Note: I would like to thank John Antonakis, Thomas Fischer, Fabiola Gerpott, Andy Loignon, Haley Woznjy, and Jill Yavorsky for suggestions related to this work.
* Address: George Banks, 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

https://doi.org/10.1016/j.leaqua.2023.101710
Available online 2 June 20231048-9843/

leadership scholarship are countless as they span micro and macro areas, many scientific disciplines, and ultimately all corners of society. In other words, the grand challenges of other areas of science and society at large are the grand challenges of leadership scholars.

Without leadership, the accomplishment of grand challenges is not possible. Thus, our current grand challenges range from creating purposeful workplaces and reducing discrimination to addressing global health concerns and promoting basic human rights such as free speech and equality. One recently accomplished grand challenge was the speedy development of safe and effective vaccines for the COVID-19 virus. Critically, just as the scientific method was necessary to create vaccines, science has been and will continue to be important for accomplishing other grand challenges that fall under the purview of leadership (Wulff et al., 2023).

We must first solve scientific puzzles in order to be able to advance theory and make meaningful contributions to practice and policymaking. Societal grand challenges are distinct from the scientific puzzles leadership scholars face. Goal-setting theory has long demonstrated that vague, lofty and ambiguous goals are harder to accomplish than those which are specific and measurable (Locke & Latham, 2002). Here is a practical example of this theory: The 23 problems or puzzles introduced by David Hilbert in the early 1900s are one of the most famous sets of specific goals in the last century (Hilbert, 1902). This example has become known to history because of the effects of stimulating collaboration and creativity, focused on a common set of problems (Fang & Casadevall, 2015; Varmus et al., 2003). Scientific puzzles are the means to develop evidence-based knowledge for practice and policymaking.

Consequently, the focus of this editorial is on the scientific puzzles faced by leadership scholars that will ultimately contribute toward the accomplishment of societal grand challenges. Science involves the systematic study of phenomena through inductive and deductive means (McDermott, 2023). The scientific method requires that findings are reproducible (e.g., one can obtain the same result with the same data and analysis; Hardwicke et al., 2018) and attempts at replication are possible (Aguinis & Solarino, 2019; Collaboration, 2015; Maxwell, Lau, & Howard, 2015). It is through the application of the scientific method that we can solve scientific puzzles in leadership.

Scientific puzzles in leadership

In this section, I now propose eight puzzles to scientists who study leadership. This list is by no means an exhaustive list. Rather, it is meant as an illustrative starting point as others will no doubt be able to propose other scientific puzzles (or refine these) that must be solved in order to accomplish broad grand challenges faced by society. In Table 1, each puzzle is presented along with a description and a set of criteria for how we might evaluate when each puzzle has been solved. It should be noted that no one study is able to solve any of these puzzles. Rather, the solutions will come from a collaborative program of work by leadership scholars. Moreover, while I attempt to present evidence of when these puzzles are solved, unlike mathematicians, we are scientifically investigating a complex and changing world. What is more likely, is that the puzzles, when solved, present relatively stable, but not final body of evidence.

#1: What are effect size benchmarks for leadership research?

Statistical benchmarks have been in existence for decades, such as the correlation and mean-difference benchmarks proposed by Jacob Cohen (1962) and later meta-analytically derived updates have been suggested by others (e.g., Bosco, Aguinis, Singh, Field, & Pierce, 2015; Gignac & Szodorai, 2016). Yet, such benchmarks typically only capture (1) covariation and not (2) temporal precedence and do not (3) eliminate alternate explanations (e.g., omitted variables; simultaneity; Hill, Johnson, Greco, O'Boyle, & Walter, 2021). For effect size benchmarks to be called such, they must reflect an actual causal mechanism, that is how a change in one variable alters another ($x_1 \rightarrow y_1$). Evidence for causal inference is a matter of degree. That said, effect size benchmarks have to be causally identified to merit the use of the term "effect size." However, this term is used loosely in our field and gives findings a veneer of policy relevance.

Because proper causal identification is key and sufficient to guide an a priori power analysis, inform Bayesian priors, or direct theory development, even one causally identified effect size estimate is better than meta-analytic estimates that are confounded. At least two effect sizes are necessary to conduct a meta-analysis. The more effect sizes we can meta-analyze, the more we can reduce random-sampling error. Furthermore, additional data means we can account for moderators that allow for understanding true variation as contingencies matter when interpreting effect size magnitudes (Kelley & Preacher, 2012). Context is important to consider in terms of the interpretation of effect sizes as small, medium, and large (Aguinis et al., 2010). For example, $\rho = .08$ may be a considered as a small effect size in a personnel section context when evaluating a selection tool. However, this same effect size might be considered large when considering the effect of an independent variable, such as executive narcissism and firm financial performance or ethical behavior and follower health.

How will we know when this puzzle is solved?

We can say that this puzzle is solved for a relationship between an independent and a dependent variable when we have effect size benchmarks that are (1) causally identified and (2) have been meta-analyzed to reduce random-sampling error and other sources of artifactual variance (e.g., measurement error). Causal identification can be achieved in a number of ways. As long as demand effects are eliminated and choices are consequential and appropriate for the given information environment (Lonati, Quiroga, Zehnder, & Antonakis, 2018), experiments are the gold standard for causal identification given the ability to leverage random assignment as well as to manipulate variables of interest (Podsakoff & Podsakoff, 2019).

However, if the key independent variable is observed (and not randomized) there are a number of other tools available to leadership scholars, such as the use of instrumental variables in two-stage least squares (2SLS) models (Bastardoz et al., 2023; Sajons, 2020) as well as quasi-experimental designs including the use of regression discontinuity, propensity score matching, selection (or treatment effects) models, and difference-in-difference (Antonakis, Bendahan, Jacquart, & Lalive, 2010; Hill et al., 2021) to establish causality in observational research. We need to be able to cumulate such evidence as well in some manner. Effect sizes can only be informative if multivariate correlations with other variables are controlled for. For instance, reporting the meta-analytic correlation of a specific personality factor with job performance cannot inform policy if the correlations of the rest of the personality factors are not first partialled out.

To summarize, establishing and refining causally identified effect size benchmarks will always be a revolving process, especially as we build and test new theory. However, as of the time of this writing, we have only limited meta-analytic effect size benchmarks that are causally identified in leadership science. For example, we have the beginning of such benchmarks in the area of charismatic leadership (Ernst et al., 2022). A great deal more work is needed. Prospective meta-analysis as a tool is one means to accelerate the process of establishing such effect size estimates as will later be discussed.

#2: What are specific leader behavioral taxonomies?

Behaviors can be defined 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, Lidicker, &

Table 1
Eight puzzles of leadership science.

Puzzles	Description	How will we know when this puzzle is solved?
1. What are effect size benchmarks for leadership research?	▪ "Effect size" benchmarks have in existence for decades, such as correlation benchmarks. Yet, such benchmarks are not effect sizes as they typically only capture (1) covariation and not (2) temporal precedence and especially not (3) a reduction of alternative explanations (e.g., omitted variables; simultaneity), etc. ▪ For leadership effect size benchmarks to be called such, they must reflect an actual causal effect (x_1 to y_1). That is, benchmarks that have been causally identified.	▪ When we have relations that are causally identified and have been meta-analyzed to reduce random-sampling error and other sources of artifactual variance ▪ When contextual moderators are accounted for that allow for an understanding of true variation.
2. What are specific leader behavioral taxonomies?	▪ The field of leadership hardly ever studies leader or follower behavior. ▪ When behavior is studied, it is in isolation which does not represent how followers simultaneously experience leader behaviors ▪ Taxonomies of leader behaviors should be formed that cluster verbal and non-verbal leader behavior at a micro-level as well as leader behavior at a more macro-level (e.g., CEO behavior)	▪ When we have specific behavioral taxonomies that can be used to train machine learning algorithms on a variety of categories of leader behaviors ▪ When we have categories of leader behaviors that can be manipulated in experiments (and behaviors not manipulated can be held constant)
3. Where do leader behaviors overlap?	▪ We must gain an understanding of how followers simultaneously experience various types of leader behaviors enacted rather than almost always studying a leader behavior in isolation or asking followers to evaluate a leader on a given style in isolation. ▪ This allows for understanding when there is true overlap of behavioral displays (the behaviors co-occur) or there is some element of redundancy.	▪ When we have behavioral evidence that demonstrates where a specified behavior does or does not exist in the presence of another. ▪ A trained algorithm could be used to score naturally occurring verbal leader behavior through sources of data such as speeches, email and social media posts as well as videos and transcripts of meetings.
4. To what extent does the effectiveness of leader behavior generalize across cultures?	▪ The majority of leadership research has been conducted in settings that are Western, Educated, Industrialized, Rich, and Democratic (WEIRD). ▪ Some leadership evidence likely generalizes to non-WEIRD contexts. However, there are also likely a number of ways existing or new knowledge will not generalize which we must understand.	▪ When we can make comparisons of causally identified effect size magnitudes in different cultures ▪ When contextual moderators are accounted for that allow for an understanding of true variation.
5. To what extent does the effectiveness of leader behavior generalize across demographic groups?	▪ With the majority of leadership research being based on surveys/questionnaires, we primarily have data on followers' evaluations of leaders (y_1). Consequently, we do not have knowledge of the extent to which certain behavioral approaches cause leader emergence and effectiveness for men relative to women. ▪ The same is true regarding other demographic groups, such as based on race and age.	▪ Leadership scientists must measure actual behavior of men and women enacted in a laboratory or field setting to understand (a) if men and women use certain behaviors at similar rates; (b) if they are evaluated differently for using the exact same behavior or if the effectiveness of the behavior differs. ▪ This same strategy must be equally explored in other demographic categories
6. To what extent do predictors, such as follower behaviors or contextual factors cause leader behaviors?	▪ Leadership is a social influence process, yet the majority if leadership models (a) focus on leaders (not followers) and (b) do not capture a dynamic exchange between "leaders" and "followers" which show the complex interplay of social influence back and forth. ▪ For instance, followers' behavior or organizational factors might proceed and even cause the behavior of the individual designated as the leader. ▪ Leader behavior is often not exogeneous, but endogenous	▪ When leadership models are developed and validate which account for a dynamic influence process.
7. What barriers exist for women in rising to leadership roles and being effective once they occupy these positions?	▪ Currently women are underrepresented for many upper level leadership positions in modern organizations. Bottom-up solutions *and* top-down solutions must be simultaneously employed to address such a matter. ▪ We must definitely establish what structural inequalities exist as barriers and remove them.	▪ When we have an increase in articles uncover bias ▪ When we have an increase of articles documenting what happens when women get power ▪ When we have additional research on interventions to reduce structural inequalities
8. What is the unifying theory of leadership?	▪ Leadership science is plagued by theory proliferation to the point that individual concepts of leadership (e.g., transformational leadership) are said to be their own theory. ▪ Like other areas of the social sciences, we lack any strong, unifying theory such as is observed in other areas of science (e.g., physics).	▪ This will be accomplished when a number of leadership theories have been properly tested and failed, retired, or combined with other leadership theories representing at least a 90% reduction in existing leadership theories. ▪ The emergence of 2-3 testable leadership theories that can explain a meaningful amount of phenomena in organizations.

Freund, 2009; p. 103). Social science research rarely studies behavior (Baumeister, Vohs, & Funder, 2007; Fischer, Hambrick, Sajons, & Quaquebeke Van, 2020). Specifically, in the area of leadership and organizational behavior, evidence indicates that only approximately 19% of studies measure actual behavior and only about 3% of variables are behavioral in nature (for a systematic review see Banks, Woznjy, & Mansfield, in press). The current state of the science means that both leader behavior (i.e., behaviors that create social influence) and follower behavior are missing from the study of leadership processes (for visual depictions see figures in Banks et al., 2023; Banks et al., in press).

What is more commonly captured than behaviors are follower evaluations of formal leaders' behaviors (y_1) as well as follower perceptions of their own attitudes (y_2) or leader evaluations of follower behavior (y_3). The folly then frequently made in leadership research, including articles published in very elite journals, is the idea that leader behavior is synonymous with follower evaluations ($x_1 = y_1$). This is a theoretical fallacy that goes unrecognized by a large number of

leadership scholars (Fischer & Sitkin, in press; Hemshorn de Sanchez, Gerpott, & Lehmann-Willenbrock, 2022). Rather, follower evaluations capture subjective assessments of leader behavior, contextual information (e.g., a leader's gender; a leader's reputation as successful or unsuccessful), traits and cues (e.g., appearance, personality intelligence) as well as bias among other things. Leader evaluations which tap into episodic rather than semantic memory, are likely to be closer to capturing actual behavior (Hansbrough et al., 2021). I, myself, have made the mistake of conflating leader behavior and follower evaluations in past research (Banks, Gooty, Ross, Williams, & Harrington, 2018; Banks, McCauley, Gardner, & Guler, 2016). Nevertheless, as leadership scientists, we must recognize our mistakes, learn from them, and continuously strive to improve.

The above critique is not to say that leader behavior has never been studied. Scholars have studied verbal behaviors (MacLaren et al., 2020; Truninger, Ruderman, Clerkin, Fernandez, & Cancro, 2021) as well as nonverbal behavior, such as eye gaze (Cheng et al., 2022; Maran, Furtner, Liegl, Kraus, & Sachse, 2019) and hand gestures (Clarke, Llewellyn, Cornelissen, & Viney, 2021). The problem is that in our top journals such behavioral studies are grossly unrepresented relative to questionnaire studies, which typically suffer from theoretical misspecification (Fischer et al., 2020) and endogeneity problems due to conflation (and endogeneity bias). Behavioral taxonomies have been introduced before. Yet, past taxonomies typically conflate behaviors with evaluations (e.g., Yukl, 2012). What is missing are non-conflated taxonomies of leader behaviors. Further, multiple taxonomies of leader behaviors should be formed that cluster verbal and non-verbal leader behavior at a micro-level as well as leader behavior at a more macro-level (e.g., executive behavior which would need a clean delineation from firm behavior).

There will always be value in using questionnaires to capture dependent variables like subjective perspectives of followers regarding their leaders. However, such measures do not capture leader behavior, and should almost never be modeled as independent variables because it engenders an intractable endogeneity problem (Fischer et al., 2020). We must make more of an effort to study leader behavior itself.

How will we know when this puzzle is solved?

Behavioral taxonomies can be created inductively using reproducible and objectively coded qualitative techniques (e.g., topic modeling, or other machine-learning techniques). Such taxonomies can also be developed deductively using the existing knowledge base (or by using a combination of inductive and deductive approaches). We can say that this puzzle has been definitively solved when we have specific behavioral taxonomies that can be used to train algorithms on a variety of categories of leader behaviors along with supporting reliability and validity evidence. Taxonomies will need to be refined and will change over time, but they should be relatively stable from year to year. Contextual moderators (e.g., in-person vs. virtual contexts) likely plays a role in their stability.

These same taxonomies can be used to guide experimental and quasi-experimental designs. Critically, such taxonomies must have at least three characteristics (Fischer & Sitkin, in press). First, they must not be limited to a positive or negative valence that may be evaluate one way or another as a function of unobserved preferences in the eye of the beholder. We must separate the behavior from how a follower might evaluate it as a good behavior or a bad behavior. As scientists, we must acknowledge this point in our behavioral taxonomies and not assign positive or negative valences. Second, behavioral taxonomies may be formed into specific taxonomies or *leadership styles*. However, such taxonomies are an oversimplification of leadership "in the wild" and ideally, we must study as many relevant behaviors as possible (or at least control for them, assuming they are objectively measured and exogenous).

Finally, any behavioral taxonomy must take into consideration how social influence is created, which is a necessary condition of characterizing the behavior as leadership. This last condition may be hardest to determine. For example, when is a behavior a counterproductive work behavior (Marcus, Taylor, Hastings, Sturm, & Weigelt, 2016) and not abusive supervision (Fischer, Tian, Lee, & Hughes, 2021)? The former is typically assumed to be a follower behavior and the latter a behavior of a leader. However, in reality, a "follower" can enact detrimental behaviors that influence a formal "leader." Of course, we must avoid a tautological bias here where concepts are defined by their outcomes (MacKenzie, 2003). That is, defining a behavior as a leader behavior simply because it creates social influence.

#3: Where do leader behaviors overlap?

A third scientific puzzle naturally follows from the second. Where do leader behaviors overlap? There could be true overlap of behavioral displays (the behaviors co-occur) or there is some element of redundancy. Thus, this puzzle closely relates to a longstanding question in leadership research around concept redundancy (for a meta-analytic review Banks et al., 2018). Once we have begun to study actual behavior in greater earnest, we may begin to understand the full extent of redundancy in the leadership literature as well as when leader behaviors truly co-occur (i.e., they co-occur but are truly not redundant). Part of the reason why redundancy concerns have not been further advanced despite many articles on the topic is the fact that scholars are continuously incentivized to introduce new concepts without intentionally positioning the new concept relative to existing ones (Newman, Harrison, Carpenter, & Rariden, 2016). Separating behavior from evaluations of that same behavior (see Puzzle #2) is a needed first step that then allows for investigating redundancy and true overlap. Many current behavior classifications are limited by concept redundancy due to overlapping positive evaluations (Fischer & Sitkin, in press). If so, separating behaviors from evaluations reduces this overlap and thus helps reduce redundancy.

As one example, consider the study of charismatic and transformational leadership which has improved greatly since earlier critiques (Van Knippenberg & Sitkin, 2013). Yet, more work could be done to examine any redundancy in behavioral measures as well as the true co-occurrence of behaviors. For instance, a story or anecdote delivered as a verbal behavior, currently classified as a charismatic leadership tactic (CLT; Antonakis, d'Adda, Weber, & Zehnder, 2022), may also be used as a transformational behavior to change the perspective of a follower (Stock, Banks, Voss, Tonidandel, & Woznyj, 2022). Or a rhetorical question (a CLT) could be used to deliver a harmful message (a destructive behavior). Consequently, one could experience two or more verbal or non-verbal behaviors at the same time which may overlap, but are not redundant.

How will we know when this puzzle is solved?

We can say that this puzzle has been definitively solved when we have behavioral evidence that demonstrates where a specified behavior does or does not exist in the presence of another. As one specific example, a trained algorithm could be used to score naturally occurring verbal leader behavior through sources of data such as speeches, email and social media posts as well as videos and transcripts of meetings (e.g., Jensen et al., 2023). Further, we would need to be able to develop AI that synthesizes *multiple* taxonomies in meaningful ways. Being able to teach an AI model to measure overlapping behavior indicates that we ourselves can understand when the behavior overlaps or is simply redundant.

We can also conduct robust qualitative studies (see McDermott, 2022) with counterfactuals as well as experimental and quasi-experimental designs (see Sieweke & Santoni, 2020) where more than one behavior is manipulated or observed in the presence of another. One of the best ways to understand something is to try to change it (a saying attributed to Kurt Lewin; Antonakis, Fenley, & Liechti, 2011). Consequently, the concerns about construct redundancy in

the leadership literature could be further informed by designs that manipulate or change the phenomenon being studied. Unfortunately, questionnaires will be very limited in solving this puzzle because followers often resort to global evaluations of leaders. This reason partly explains why some have wondered, "is leadership more than "I like my boss?" (Yammarino, Cheong, Kim, & Tsai, 2020; p. 1). In sum, we must gain an understanding of how followers experience various types of leader behaviors enacted simultaneously rather than almost always studying a leader behavior in isolation or asking followers to evaluate a leader on a given style in isolation.

#4: To what extent does the effectiveness of leader behavior generalize across cultures?

Puzzles #2 and #3 propose establishing behavioral taxonomies and considering how behaviors co-occur. Out of necessity, we must also consider our leadership scholarship in nontraditional settings. In other words, settings that are not Western, Educated, Industrialized, Rich, and Democratic that is, WEIRD (Arnett, 2008; Henrich, Heine, & Norenzayan, 2010; Rad, Martingano, & Ginges, 2018). Almost 100% of psychological research comes from about 10% of the world's population presenting a potential bias if such research is attempted to be generalized. Leadership research is not unique compared to general psychology. In many ways, leadership scholarship likely generalizes to non-WEIRD contexts. However, there are also likely a number of ways existing or new knowledge will not generalize.

How will we know when this puzzle is solved?

We can say that this puzzle has been definitively solved when causally identified meta-analytic reviews are completed and which are able to include non-WEIRD samples to the extent that such a knowledge base can be considered as a robust moderator. Thus, this puzzle will be largely resolved when we are able to make comparisons of causally identified effect size magnitudes in different cultures. Once again, prospective meta-analysis is one means to accelerate the inclusion of non-WEIRD samples as a complement to traditional retrospective meta-analyses.

There are a number of challenges faced by scholars in non-WEIRD contexts who seek to conduct leadership research. For instance, such scholars may not have the necessary training (e.g., in conducting causally identified research) or resources (e.g., basic journal access) to meet standards of journals owned by Western publishers. Alternatively, there may be bias from reviewers or editors that occurs when submitting to journals owned by Western publishers because the samples come from nonwestern contexts. Or scholars in non-WEIRD contexts may simply publish in non-Western owned journals and Western scholars may not seek out journals not published, for instance, in English. One would hope that meta-analytic scholars who conduct proper systematic reviews would not restrict their searches to English only terms, but some do. Most search engines of databases are capable of aiding in this process. Consequently, to address this puzzle, we must ensure that non-WEIRD samples exist (we seek ways to help cultivate such samples through collaborations and grants) and that we make efforts to reduce bias in publishing and/or seek out such samples where they are published.

#5: To what extent does the effectiveness of leader behavior generalize across demographic groups?

Related to the previous puzzle, as we progress in the discussion of leadership science as puzzles, we can begin to imagine what other puzzle pieces must be filled in for us to have a more complete picture of the phenomenon of interest. Naturally, another piece we must add is a true understanding of the extent to which the effectiveness of leader behavior generalizes across demographic groups. Here, I refer to effectiveness broadly to include behaviors which help one emerge as a leader and to be successful once one is in a leadership position. As previously noted, the majority of leadership research is based on questionnaires (Fischer et al., 2020). Consequently, we have some knowledge of how followers evaluate their leaders, but not what behaviors are enacted to cause those evaluations. For example, meta-analytic evidence shows there is almost no difference between how men and women as leaders are evaluated by followers on charisma (the 95% confidence interval incudes 0; Banks et al., 2017); though the usual caveats should be considered given that this research was based on questionnaires and the behaviors were not exogenously manipulated across men and women (and then evaluated). We do not know to what extent women and men enact the same behaviors to cause evaluations of charisma. Or to what extent women and men enact different (sets of) behaviors. To what extent are they evaluated the same for using the same behaviors?

Again, it is quite possible that there are times men and women are evaluated the same by followers for using the same behavior. However, there are also potentially times that men and women are evaluated differently by followers based on societal expectations and role congruity. As one example, if men leaders dominate the speaking time during a meeting, they may be evaluated more favorably than women leaders who speak for the same length of time (MacLaren et al., 2020). The same result holds true for race and other demographic categories in terms of the need to understand how leader behavior effectiveness generalizes.

How will we know when this puzzle is solved?

We can say that this puzzle has been definitively solved when meta-analytic reviews, using clean causal data, are able to compare and model one or more demographic categories simultaneously. There are a number of moderating factors that would have to be accounted for in the meta-analytic reviews, such as intersectionality (e.g., between race and gender). Or such studies might consider how leaders adapt their behaviors based on formal/informal roles, level of the leader, responses from followers (see Puzzle #6). Study preregistration and/or registered reports (to be discussed later) are useful tools in answering any of the scientific puzzles (for reviews see Chambers & Tzavella, 2022; Toth et al., 2021). However, they may be especially valuable in this context to avoid publication bias. For instance, if findings generally show evidence for gender differences on a particular leadership outcome, it may be harder to publish a study showing null results because this result may be inconsistent with the expectations of some reviewers. Else, if results confirm a gender difference, reviewers may question the novelty of the finding.

When comparing men and women in leadership research it is important that we publish the results regardless of the findings, which is a key value of this journal. For instance, recent findings have questioned whether the Queen Bee (Arvate, Galilea, & Todescat, 2018) and the Glass Cliff phenomena are simply myths (Bechtoldt, Bannier, & Rock, 2019)—such research would be difficult to publish in journals that require novelty, statistically significant results, and a theoretical contribution. Others have published mixed and inconclusive findings on the benefits of corporate board diversity (Zattoni, Leventis, Van Ees, & De Masi, 2022) as well as conflicting findings on the persistence of racial stereotypes (Petsko & Rosette, 2022; Ubaka, Lu, & Gutierrez, 2022). In sum, we must understand the effectiveness of leader behavior across demographic groups using a combination of rigorous methodologies and then publish the results regardless of what we find.

#6: To what extent do predictors, such as leader individual differences, follower behaviors or contextual factors, cause leader behaviors?

Leadership is a social influence process, which in reality has a dynamic back and forth element between the influencer and the target of the influence (Heggestad et al., 2023). Yet the majority of leadership models (1) focus on leaders (not followers) (Bastardoz & Van

Vugt, 2019) and (2) do not capture a dynamic exchange between "leaders" and "followers" which show the complex interplay of social influence back and forth (Oc & Bashshur, 2013). For instance, followers' behavior (e.g., words or emotional displays) or contextual factors might proceed and even cause the behavior of the individual designated as the leader. Thus, observed (i.e., measured) leader behavior is often endogenous and not exogenous (Güntner, Klonek, Lehmann-Willenbrock, & Kauffeld, 2020). Consequently, we must develop and validate leadership models that illustrate the extent to which predictors, such as follower behaviors or contextual factors, cause leader behaviors. Of course, with exogenous manipulations of leader behavior, we can safely estimate the one-way effect of leader behavior on follower outcomes. Related to this point, we do not know enough yet regarding biological causes of leader emergence and effectiveness, such as genes (Van Vugt & von Rueden, 2020) or stable individual differences (Antonakis, Day, & Schyns, 2012).

How will we know when this puzzle is solved?

We can say that this puzzle has been definitively solved when leadership models are developed which account for a dynamic influence process. This goal requires theorizing and methodological designs that account for time (Aguinis & Bakker, 2021; Shipp & Cole, 2015). However, simply because leader behavior is measured at Time 1 and a follower outcome is measured at Time 2, we cannot assume the leader behavior to be exogenous (Güntner et al., 2020). That is, just because we measure one variable (x_1) before the other (y_1), we cannot theoretically or empirically assume that $x_1 \rightarrow y_1$. In fact, the sequence may be $y_1 \rightarrow x_1 \rightarrow y_2$; omitted causes could explain the relations too over time. Whereas experiments are useful at determining causal inferences, many typical experimental designs are limited in capturing dynamic processes. To best solve this puzzle, a triangulation approach is required that uses a mixture of methodologies in field and laboratory settings and at the same time ensures correct causal specification in models.

#7: What barriers exist for women in rising to leadership roles and being effective once they occupy these positions?

Currently women are underrepresented for many upper level leadership positions in modern organizations (Paustian-Underdahl, Walker, & Woehr, 2014). It has been reported that *less than 5%* of studies in relevant journals focus on specific interventions for addressing gender inequality (Lau, Scott, Warren, & Bligh, in press). Bottom-up solutions *and* top-down solutions must be simultaneously employed to address such a matter (Perez, 2019). Puzzle #5 includes a focus on such bottom-up solutions to advance gender equality through leadership training and development. However, we must also consider structural inequalities in the form of our organizational processes that produce barriers for women to be successful as leaders (Martell, Lane, & Emrich, 1996; Samuelson, Levine, Barth, Wessel, & Grand, 2019). In many ways, we already have a grasp of the obstacles that women face, such as systematic bias in the workplace (Eagly & Karau, 2002), social challenges including caregiving (Yavorsky, Kamp Dush, & Schoppe-Sullivan, 2015), and the devaluation of women dominated jobs (Yavorsky & Dill, 2020).

What is required, is actually a puzzle within a puzzle to more explicitly understand how macro challenges may be addressed with "small wins" (Correll, 2017). Similar to Lau and colleagues (in press), I would suggest that more research is needed to focus on solutions in the leadership literature that can lead to policymaking. This work also partly necessitates understanding privileges or advantages that men face (i.e., the lack of barriers).

How will we know when this puzzle is solved?

We can say that this puzzle has been solved when we have an increase in articles uncovering bias (Treviño, Gomez-Mejia, Balkin, & Mixon, 2018) as well as an increase in articles documenting what happens when women exogenously get power, especially in field settings (Arvate et al., 2018; Matsa & Miller, 2013). Further, additional research is required on the impact of diversity outcomes as well as on how interventions can reduce bias against women that manifest from structural inequalities. Of course, similar efforts to solve this puzzle can be applied to inequality more broadly (e.g., sexual identity, nationality, ethnicity, age, social class) as well as to better understanding intersectionality.

#8: What is the unifying theory of leadership?

Leadership science is plagued by theory proliferation to the point that individual concepts of leadership (e.g., transformational leadership) and in particular their measures are said to be their own theory (Van Knippenberg & Sitkin, 2013). In fact, individual leadership concepts are essentially characters in a much broader theoretical story (Shepherd & Suddaby, 2017). Unlike other areas of the social sciences (e.g., economics, Wulff et al., 2023), we lack strong, unifying theoretical frameworks (Pfeffer, 1993). Solving the puzzles that precede this last puzzle will be necessary to tell the story of the critical phenomena in the leadership domain. Paradigms have a number of helpful benefits which include aiding in the coordination of efforts, attracting resources for our discipline, and having more effective stakeholder communication.

How will we know when this puzzle is solved?

We can say that this puzzle has been definitively solved when a number of leadership theories have been properly tested and failed, retired, or combined with other leadership theories representing at least a 90% reduction in existing "leadership theories." Most tests of theory lack theoretical precision (Edwards & Berry, 2010). The emergence of 2–3 testable leadership theories that can explain a meaningful amount of phenomena in organizations is required.

Four actionable guidelines for accomplishing our scientific puzzles

The puzzles presented above are specific yet rather ambitious and will be no easy undertaking to answer. Hence, we must now consider actionable steps for accomplishing these puzzles. These actionable guidelines also serve to address the scientific diseases previously noted in *The Leadership Quarterly* including (1) shunning null results and publishing only statistically significant results (significosis), (2) over emphasis on novelty (neophilia) and (3) fetishizing new theory (theorrhea), (4) poor methodological rigor (arigorium), and (5) a disjointed scholarship (disjunctivitis) (Antonakis, 2017b). The action guidelines are summarized in Table 2 along with supporting descriptions.

Open science

First, open science practices should be leveraged as appropriate given the design and purpose of the study. Open science represents a broad set of constellation of values, categories of activities, and very specific practices (Banks et al., 2019; Nosek & Bar-Anan, 2012; Nosek, Spies, & Motyl, 2012). The broad nature of open science as a concept can lead to some confusion over what open science represents. At *The Leadership Quarterly*, we view open science as a set of values related to those identified by the National Academies of Science, Engineering, and Medicine (2017): objectivity, honesty, openness, accountability, fairness, and stewardship. Castille and colleagues (2022) suggested that there are a buffet of open science practices that scientists can choose from and showed how different practices can be enacted to be consistent with these six values. Similarly, I encourage scientists who wish to help answer scientific puzzles to consider these

Table 2
Actionable guidelines for solving leadership puzzles.

Actionable guidelines	Description	Examples
1. Open science	■ Open science practices should be leveraged as is appropriate given the design and purpose of study.	■ Study preregistration (i.e., explicit, documented planning) ■ Open data, analytic code, study materials ■ Registered reports and results blind reviews ■ Transparent reporting of authorship practices ■ Accepting/publishing/embracing null findings
2. Interdisciplinary collaboration	■ Large scale, interdisciplinary collaborations need to be used to accelerate the advancement of science.	■ Prospective meta-analysis ■ Adversarial collaborations on direct and conceptual replications
3. Triangulation	■ Multiple methodologies can be used to study the same phenomena from different perspectives.	■ Inductive, deductive, and abductive qualitative research ■ Inductive, deductive, and abductive quantitative research ■ Experimental (in field or laboratory settings) ■ Quasi-experimental techniques (e.g., regression discontinuity or difference-in-difference designs) ■ Survey-based approaches ■ Data science (e.g., topic modeling, deep learning models).
4. Fixing common methodological mistakes	■ Common methodological mistakes must be minimized in quantitative and qualitative research in the theorizing, design, analysis, and interpretation phases	■ Comparing "poison" vs. "medicine" groups in experiments ■ Sampling on the dependent variable ■ Conducting mediation tests with endogenous variables as the independent variable

values and then to align their research with these values as it makes sense.

As one example, consider study preregistration. Study preregistration can best be thought of as a plan. Regardless of one's study design and research questions of interest, it is a good idea to begin with a documented plan even if that plan changes (such changes can be further documented). In other words, one may be conducting qualitative research (e.g., Qualitative Content Analysis; QCA), gathering archival data, or planning an experiment. Whether the work is inductive, deductive, or abductive, some basic plan at the start is beneficial for the research team and future scholars who wish to build upon their work. Still, some scholars may identify legitimate reasons why a preregistration does not make sense for a particular study. Instead, those scholars can share other materials, such as interview and focus group guides or data, analytic code, and a transparency checklist (Aczel et al., 2020).

As a second example, a registered report is the submission of an Introduction and Methods section to a journal. It helps to give researchers the opportunity for early feedback from reviewers (before a study is executed) and hopefully an in-principle acceptance (IPA) is awarded meaning the work will be published regardless of the results as long as researchers stick to their study plan. Despite the benefits of registered reports, they may not be ideal for all types of studies. During March of 2020 when the COVID-19 pandemic was unfolding, scholars would not have had time to complete a registered report submission and multiple rounds of revision before receiving an IPA. Hence, other open science practices, such as study preregistration followed by a results-blind review, may have better served as a means to enact values. Thus, whereas specific open science practices may be encouraged, what is most important is that our scientific community remember our shared values around openness and collaboration and to enact those values in our daily work. Without open science, we will not solve the scientific puzzles described above.

Interdisciplinary collaboration

Second, and very closely related to open science is interdisciplinary collaboration. Some have asked, "is competition ruining science?" (Fang & Casadevall, 2015). I think the answer is yes and no. In the rapid development of the COVID-19 vaccines, we saw collaborative competition successfully at work (Dolgin, 2021). There will always be benefits to individual and small team research because creativity, recognition and competition for resources can spur innovation. However, there is also much to be gained from large-scale collaboration. Collaboration can also take many forms, such as adversarial collaborations on direct and conceptual replications. This endeavor might involve, for instance, prospective meta-analysis in which a very large team of collaborators works to conduct novel research as well as conduct exact or conceptual replications (for a tutorial see Seidler et al., 2019). Scientific competition will always be present because of the drive for evidence-based advantages. What is required in the 21st century though is an increase in interdisciplinary collaboration.

Triangulation using different methodologies

Third, the solutions to leadership science puzzles will require multiple methodologies. Every methodology has its advantages and disadvantages. No one study or its design is perfect. That is, there are always limitations in study designs. Consequently, we can better solve our puzzles by leveraging triangulation which is defined as the use of "multiple reference points to locate an object's exact position" (Jick, 1979; p. 602). Methods can include qualitative (e.g., inductive or deductive), experimental (in field or laboratory settings) and quasi-experimental techniques (e.g., regression discontinuity or difference-in-difference designs) as well as survey-based approaches and data science (e.g., topic modeling, deep learning models). Each of these techniques has its advantages and disadvantages. What is most important to us is that the design is robust, with strong counterfactual evidence, and that the findings can be reproduced and replications attempted, which is a hallmark of the scientific method. By combining techniques within or across a set of studies, we can build a stronger body of evidence to more definitely answer scientific puzzles.

Fixing common methodological mistakes

As a fourth and final actionable guideline, common methodological mistakes must be cleaned up. These mistakes are relatively easy to

minimize in quantitative and qualitative research in design, analysis, and interpretation phases of the scientific process. For a list of common methodological mistakes see Wulff et al. (2023).

The role of *The Leadership Quarterly*

This journal will contribute toward the pursuit of the puzzles listed above by publishing meaningful and rigorous scientific research. This includes scientific work that contributes to theory, practice, policy-making, and research methods. Consistent with the previous editorial team, we will continue to solicit submissions from disciplines in the social and natural sciences. Because of this broad multidisciplinary focus for the journal, the editorial team of associate editors and the editorial board are quite diverse in terms of their content and methodological expertise (https://www.sciencedirect.com/journal/the-leadership-quarterly/about/editorial-board).

As in past years, the journal will continue to accept full-length articles (e.g., deductive, inductive, and abductive studies; theoretical articles, systematic, narrative, and meta-analytic reviews; critiques or reflection pieces; applied papers and methodological articles), short communications (e.g., replication articles, commentaries on published articles) and proposals which are either registered reports or results-masked articles). There will be yearly review articles to be handled via a two-stage review process. For a full description of each type of submission, please visit the journal's website (https://www.sciencedirect.com/journal/the-leadership-quarterly).

Finally, *The Leadership Quarterly* also solicits prospective meta-analyses (Seidler et al., 2019), ideally as a part of the registered report process. As an illustrative example, in a previous meta-analysis by Banks et al. (2017), a number of limitations in the primary study literature of leadership were recognized. Rather than wait 10-20 years in the hopes that the limitations would be addressed, Ernst et al. (2022) proceeded to complete a set of preregistered experiments via an international collaboration. I encourage leadership scholars to consider a similar type of approach. Submitting such a proposal via a registered report reduces the risk involved and could aid in recruiting collaborators after the IPA.

We still welcome null results studies; all findings will be held to the same standard of methodological rigor (for reviews of our methodological standards please carefully read our author guidelines and the following editorials: Antonakis, 2017a, 2017b; Antonakis et al., 2019; Wulff et al., 2023). Associate editors will continue to decision manuscripts independently of the Editor (who assigns the reviewers) on the basis of the journal's editorial policies (Antonakis, 2017a, 2017b; Antonakis et al., 2019; Wulff et al., 2023). Former editors or distinguished scholars in the field will be selected to decision manuscripts of the members of the current associate editor team with the spirit of reducing conflicts of interest. The reviewers' comments serve as input for decisions and "vote counting" is not used. Reviewer training at the journal is targeted to ensure that we not only have experts review manuscripts, but that the reviews are conducted in accordance with our vision and policies.

The journal has had (Antonakis et al., 2019) and will continue to have a high desk rejection rate; the acceptance rate is still low (i.e., about 5% over the last three years). We do not have a goal rate for acceptances. Rather, the acceptance rate we have is a simple by-product of our values, bearing in mind that we have a responsibility to society to be ethical and economical. That is, if a paper is likely to be rejected after a full review because of a fundamental flaw, we see the value in a rejection after seven days compared to 90 days. In this way, authors still get feedback on the fundamental flaw, but do not have to wait as long for a decision. Further, reviewer and action editor time are preserved for working on manuscripts with a higher likelihood of publication. This modus operandi shifts editors' mental view of submissions. Instead of looking for flaws to justify a rejection, our editorial team can be more constructive as they are used to reading high quality work.

All of this being said, in many ways, this journal is one of the most straightforward to publish in because (1) we have high standards but (2) we are very transparent about our standards as laid out in our author guidelines (https://www.elsevier.com/journals/the-leadership-quarterly/1048-9843/guide-for-authors) and recent editorials (Antonakis, 2017a, 2017b; Antonakis et al., 2019; Wulff et al., 2023). That is, decisions are less idiosyncratic and those who wish to submit have a very clear target, albeit a high one. We think it is better to have a clear, high target than one that is ambiguous and vague. If you conduct sound scientific research with a meaningful contribution, we will publish it regardless of whether the results are statistically significant or null, counter to past research or confirm past research. Many have lamented the pressure to publish and the need to churn out journal submissions (Aguinis, Cummings, Ramani, & Cummings, 2020; Nosek & Bar-Anan, 2012). This is "fast food" science [1]. Rather, we are "in support of slow science" (Antonakis, 2023). Take your time, put your best efforts into your work, and you need not shop around such work to multiple journals in the hopes of it finding a home. We will publish it.

Conclusion

In the current editorial, I proposed a challenge to leadership scientists to answer eight puzzles. These puzzles will not be easy to solve and as noted, they require improved theorizing, empirical design, and notably, resources. We must work *together*. Whereas individual ingenuity and creativity are necessary, these puzzles cannot be deciphered without interdisciplinary collaboration and open science. Let us galvanize our energy, traverse disciplinary boundaries, and accelerate our science towards new frontiers. In this way, we might lift the veil behind which the future lies hidden, as Hilbert proposed (1902).

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

Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., ... Wagenmakers, E. J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour, 4*(1), 4–6.

Aguinis, H., & Bakker, R. M. (2021). Time is of the essence: Improving the conceptualization and measurement of time. *Human Resource Management Review, 31*(2), 100763.

Aguinis, H., Cummings, C., Ramani, R. S., & Cummings, T. G. (2020). "An A is an A": The new bottom line for valuing academic research. *Academy of Management Perspectives, 34*(1), 135–154.

Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal, 40*(8), 1291–1315.

Aguinis, H., Werner, S., Lanza Abbott, J., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods, 13*(3), 515–539.

Antonakis, J. (2017a). Editorial: The future of The Leadership Quarterly. *The Leadership Quarterly, 28*, 1–4.

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

Antonakis, J. (2023). In support of slow science: Robust, replicable, and reproducible. *The Leadership Quarterly*.

Antonakis, J., Banks, G. C., Bastardoz, N., Cole, M., Day, D., Eagly, A., ... Haslam, A. (2019). The leadership quarterly: State of the journal. *The Leadership Quarterly, 30*(1), 1–9.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086–1120.

[1] The author is unaware of the origin and a citation for this term.

Antonakis, J., d'Adda, G., Weber, R. A., & Zehnder, C. (2022). "Just words? Just speeches?" On the economic value of charismatic leadership. *Management Science, 68*, 6355–7064.
Antonakis, J., Day, D. V., & Schyns, B. (2012). Leadership and individual differences: At the cusp of a renaissance. *The Leadership Quarterly, 23*, 643–650.
Antonakis, J., Fenley, M., & Liechti, S. (2011). Can charisma be taught? Tests of two interventions. *Academy of Management Learning & Education, 10*(3), 374–396.
Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist, 63*, 602–614.
Arvate, P. R., Galilea, G. W., & Todescat, I. (2018). The queen bee: A myth? The effect of top-level female leadership on subordinate females. *The Leadership Quarterly, 29*(5), 533–548.
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., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2019). Answers to 18 questions about open science practices. *Journal of business and Psychology, 34*, 257–270.
Banks, G. C., Gooty, J., Ross, R. L., Williams, C. E., & Harrington, N. T. (2018). Construct redundancy in leader behaviors: A review and agenda for the future. *The Leadership Quarterly, 29*(1), 236–251.
Banks, G. C., McCauley, K. D., Gardner, W. L., & Guler, C. E. (2016). A meta-analytic review of authentic and transformational leadership: A test for redundancy. *The Leadership Quarterly, 27*(4), 634–652.
Banks, G. C., Pollack, J. M., Bochantin, J. E., Kirkman, B. L., Whelpley, C. E., & O'Boyle, E. H. (2016). Management's science–practice gap: A grand challenge for all stakeholders. *Academy of Management Journal, 59*(6), 2205–2231.
Banks, G. C., Woznjy, H., & Mansfield, C. (in press). Where is "behavior" in organizational behavior? A call for a revolution in leadership research and beyond. *The Leadership Quarterly*.
Banks, G. C., Ross, R., Toth, A. A., Tonidandel, S., Goloujeh, A. M., Dou, W., & Wesslen, R. (2023). The triangulation of ethical leader signals using qualitative, experimental, and data science methods. *The Leadership Quarterly, 34* 101658.
Bastardoz, N., Matthews, M., Sajons, G., Ransom, T., Kelemen, T. K., & Matthews, S. H. (2023). Instrumental Variables Estimation: Assumptions, Pitfalls, and Guidelines. *The Leadership Quarterly*. https://doi.org/10.1016/j.leaqua.2022.101673.
Bastardoz, N., & Van Vugt, M. (2019). The nature of followership: Evolutionary analysis and review. *The Leadership Quarterly, 30*(1), 81–95.
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.
Bechtoldt, M. N., Bannier, C. E., & Rock, B. (2019). The glass cliff myth?–Evidence from Germany and the UK. *The Leadership Quarterly, 30*(3), 273–297.
Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Pschology, 100*(2), 431–449.
Castille, C. M., Kreamer, L. M., Albritton, B. H., Banks, G. C., & Rogelberg, S. G. (2022). The Open Science Challenge: Adopt One Practice that Enacts Widely Shared Values. *Journal of business and Psychology*, 1–9.
Chambers, C. D., & Tzavella, L. (2022). The past, present and future of registered reports. *Nature Human Behaviour, 6*(1), 29–42.
Cheng, J. T., Gerpott, F. H., Benson, A. J., Bucker, B., Foulsham, T., Lansu, T. A., ... Tsuchiya, K. (2022). Eye gaze and visual attention as a window into leadership and followership: A review of empirical insights and future directions. *The Leadership Quarterly*, 101654.
Clarke, J. S., Llewellyn, N., Cornelissen, J., & Viney, R. (2021). Gesture analysis and organizational research: The development and application of a protocol for naturalistic settings. *Organizational Research Methods, 24*(1), 140–171.
Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145–153.
Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.
Collins, P. Y., Patel, V., Joestl, S. S., March, D., Insel, T. R., Daar, A. S., ... Fairburn, C. (2011). Grand challenges in global mental health. *Nature, 475*(7354), 27–30.
Correll, S. J. (2017). SWS 2016 Feminist Lecture: Reducing gender biases in modern workplaces: A small wins approach to organizational change. *Gender & Society, 31*(6), 725–750.
Daar, A. S., Singer, P. A., Leah Persad, D., Pramming, S. K., Matthews, D. R., Beaglehole, R., ... Ganguly, N. (2007). Grand challenges in chronic non-communicable diseases. *Nature, 450*(7169), 494–496.
Day, D. V., & Antonakis, J. (2013). The future of leadership. *The Wiley-Blackwell Handbook of the Psychology of Leadership, Change, and Organizational Development*, 221–235.
Day, D. V., Fleenor, J. W., Atwater, L. E., Sturm, R. E., & McKee, R. A. (2014). Advances in leader and leadership development: A review of 25 years of research and theory. *The Leadership Quarterly, 25*(1), 63–82.
Doepke, M., Tertilt, M., & Voena, A. (2012). The economics and politics of women's rights. *Annual Review of Economics, 4*(1), 339–372.
Dolgin, E. (2021). The tangled history of mRNA vaccines. *Nature, 597*, 318–324.
Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3), 573–598.
Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods, 13*(4), 668–689.
Ernst, B. A., Banks, G. C., Loignon, A. C., Frear, K. A., Williams, C. E., Arciniega, L. M., ... Subramanian, D. (2022). Virtual charismatic leadership and signaling theory: A prospective meta-analysis in five countries. *The Leadership Quarterly, 33*(5), 101541.
Fang, F. C., & Casadevall, A. (2015). Competitive science: is competition ruining science? In (Vol. 83, pp. 1229–1233): Am Soc Microbiol.
Fischer, T., Hambrick, D. C., Sajons, G. B., & Quaquebeke Van, N. (2020). Beyond the ritualized use of questionnaires: Toward a science of actual behaviors and psychological states. *The Leadership Quarterly, 31*, 101449.
Fischer, T., & Sitkin, S. B. (in press). Leadership styles: A comprehensive assessment and way forward. *Academy of Management Annals*.
Fischer, T., Tian, A. W., Lee, A., & Hughes, D. J. (2021). Abusive supervision: A systematic review and fundamental rethink. *The Leadership Quarterly, 32*(6), 101540.
Gardner, W. L., Lowe, K. B., Meuser, J. D., Noghani, F., Gullifor, D. P., & Cogliser, C. C. (2020). The leadership trilogy: A review of the third decade of the leadership quarterly. *The Leadership Quarterly, 31*(1), 101379.
George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and tackling societal grand challenges through management research. *Academy of Management journal, 59*(6), 1880–1895.
Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78.
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.
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.
Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., ... Henry Tessler, M. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science, 5*(8), 180448.
Heggestad, E. D., Nicole Voss, E., Toth, A. A., Ross, R. L., Banks, G. C., & Canevello, A. (2023). Two Meanings of "Social Skills": Proposing an Integrative Social Skills Framework. *Group & Organization Management*. 10596011221151006.
Hemshorn de Sanchez, C. S., Gerpott, F. H., & Lehmann-Willenbrock, N. (2022). A review and future agenda for behavioral research on leader–follower interactions at different temporal scopes. *Journal of Organizational Behavior, 43*(2), 342–368.
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). We agree it's WEIRD, but is it WEIRD enough. *Behavioral Brain Sciences, 33*(2–3), 61–83.
Hilbert, D. (1902). Mathematical Problems. *Bulletin of the American Mathematical Society, 8*, 437–479.
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*(1), 105–143.
Jensen, U., Rohner, D., Bornet, O., Carron, D., Garner, P. N., Loupi, D., & Antonakis, J. (2023). Combating COVID-19 with charisma: Evidence on governor speeches in the United States. *The Leadership Quarterly*, 101702.
Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*, 602–611.
Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological methods, 17*(2), 137.
Kintigh, K. W., Altschul, J. H., Beaudry, M. C., Drennan, R. D., Kinzig, A. P., Kohler, T. A., ... Pauketat, T. R. (2014). *Grand challenges for archaeology. American antiquity, 79*(1), 5–24.
Lau, V. W., Scott, V. L., Warren, M. A., & Bligh, M. C. (in press). Moving from problems to solutions: A review of gender equality interventions at work using an ecological systems approach. *Journal of Organizational Behavior*.
Levitis, D. A., Lidicker, W. Z., & Freund, G. (2009). Behavioural biologists do not agree on what constitutes behaviour. *Animal behaviour, 78*(1), 103–110.
Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*, 705–717.
Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management, 64*, 19–40.
MacKenzie, S. B. (2003). The dangers of poor construct conceptualization. *Journal of the Academy of Marketing Science, 31*, 323–326.
MacLaren, N. G., Yammarino, F. J., Dionne, S. D., Sayama, H., Mumford, M. D., Connelly, S., ... Kulkarni, A. (2020). Testing the babble hypothesis: Speaking time predicts leader emergence in small groups. *The Leadership Quarterly, 31*(5), 101409.
Mandela, N. (1993). South Africa's future foreign policy. *Foreign Affairs*, 86–97.
Manley, J. B., Anastas, P. T., & Cue, B. W. (2008). Frontiers in Green Chemistry: Meeting the grand challenges for sustainability in R&D and manufacturing. *Journal of Cleaner Production, 16*(6), 743–750.
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.
Marcus, B., Taylor, O. A., Hastings, S. E., Sturm, A., & Weigelt, O. (2016). The structure of counterproductive work behavior: A review, a structural meta-analysis, and a primary study. *Journal of Management, 42*(1), 203–233.
Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-female differences: A computer simulation. *American Psychologist, 51*, 157–158.
Matsa, D. A., & Miller, A. R. (2013). A female style in corporate leadership? Evidence from quotas. *American Economic Journal: Applied Economics, 5*, 136–169.
Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70*(6), 487–498.
McDermott, R. (2023). On the scientific study of small samples: Challenges confronting quantitative, qualitative and case study methodologies. *The Leadership Quarterly*, 101675.

Mertens, P., & Barbian, D. (2015). Researching "Grand Challenges" A "Grand Challenge". *Business & Information Systems Engineering, 57*, 391–403.

National Academies of Sciences Engineering and Medicine. (2017). Fostering integrity in research. National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK475953/pdf/Bookshelf_NBK475953.pdf.

Newman, D. A., Harrison, D. A., Carpenter, N. C., & Rariden, S. M. (2016). Construct mixology: Forming new management constructs by combining old ones. *Academy of Management Annals, 10*(1), 943–995.

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry, 23*(3), 217–243.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615–631.

Oc, B., & Bashshur, M. R. (2013). Followership, leadership and social influence. *The Leadership Quarterly, 24*(6), 919–934.

Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of applied psychology, 99*(6), 1129.

Perez, C. C. (2019). *Invisible women: Data bias in a world designed for men.* Abrams.

Petsko, C. D., & Rosette, A. S. (2022). Are leaders still presumed white by default? Racial bias in leader categorization revisited. *Journal of Applied Psychology*..

Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *Academy of Management Review, 18*(4), 599–620.

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.

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401–11405.

Robbins, S. P., & Judge, T. A. (2017). *Essentials of Organizational Behavior* (14th ed.). Upper Saddle River, NJ: Prentice Hall.

Sajons, G. B. (2020). Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables Retrieved from. *The Leadership Quarterly, 31*(5), 101348 http://www.sciencedirect.com/science/article/pii/S1048984318300523.

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 stratification. *The Leadership Quarterly, 30*, 101314.

Seidler, A. L., Hunter, K. E., Cheyne, S., Ghersi, D., Berlin, J. A., & Askie, L. (2019). A guide to prospective meta-analysis. *British Medical Journal, 367*.

Shepherd, D. A., & Suddaby, R. (2017). Theory building: A review and integration. *Journal of Management, 43*(1), 59–86.

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 of Organizational Psychology and Organizational Behavior, 2*(1), 237–260.

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

Stenger, R. (2001). Man on the moon: Kennedy speech ignited the dream. *CNN. com, 25*.

Stephens, J. C., Cavanaugh, M. L., Gradie, M. I., Mador, M. L., & Kidd, K. K. (1990). Mapping the human genome: Current status. *Science, 250*(4978), 237–244.

Stock, G., Banks, G. C., Voss, E. N., Tonidandel, S., & Woznyj, H. (2022). Putting leader (follower) behavior back into transformational leadership: A theoretical and empirical course correction. *The Leadership Quarterly*, 101632.

Toth, A. A., Banks, G. C., Mellor, D., O'Boyle, E. H., Dickson, A., Davis, D. J., ... Borns, J. (2021). Study preregistration: An evaluation of a method for transparent reporting. *Journal of business and Psychology, 36*, 553–571.

Treviño, L. J., Gomez-Mejia, L. R., Balkin, D. B., & Mixon, F. G. (2018). Meritocracies or masculinities? The differential allocation of named professorships by gender in the academy. *Journal of Management, 44*(3), 972–1000.

Truninger, M., Ruderman, M. N., Clerkin, C., Fernandez, K. C., & Cancro, D. (2021). Sounds like a leader: An ascription–actuality approach to examining leader emergence and effectiveness. *The Leadership Quarterly, 32*(5), 101420.

Ubaka, A., Lu, X., & Gutierrez, L. (2022). Testing the generalizability of the white leadership standard in the post-Obama era. *The Leadership Quarterly*, 101591.

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.

Van Vugt, M., & von Rueden, C. R. (2020). From genes to minds to cultures: Evolutionary approaches to leadership. *The Leadership Quarterly, 31*(2), 101404.

Varmus, H., Klausner, R., Zerhouni, E., Acharya, T., Daar, A., & Singer, P. (2003). Grand challenges in global health. In (Vol. 302, pp. 398–399): American Association for the Advancement of Science.

Wulff, J. N., Sajons, G. B., Pogrebna, G., Lonati, S., Bastardoz, N., Banks, G. C., & Antonakis, J. (2023). Common methodological mistakes. *The Leadership Quarterly, 34*.

Yammarino, F. J., Cheong, M., Kim, J., & Tsai, C.-Y. (2020). Is leadership more than "I like my boss"? *Research in personnel and human resources management* (Vol. 38, pp. 1–55). Emerald Publishing Limited.

Yavorsky, J. E., & Dill, J. (2020). Unemployment and men's entrance into female-dominated jobs. *Social Science Research, 85*, 102373.

Yavorsky, J. E., Kamp Dush, C. M., & Schoppe-Sullivan, S. (2015). The production of inequality: The gender division of labor across the transition to parenthood. *Journal of Marriage Family, 77*(3), 662–679.

Yukl, G. (2012). Effective leadership behavior: What we know and what questions need more attention. *Academy of Management Perspectives, 26*(4), 66–85.

Zattoni, A., Leventis, S., Van Ees, H., & De Masi, S. (2022). Board diversity's antecedents and consequences: A review and research agenda. *The Leadership Quarterly*, 101659.