The Leadership Quarterly 34 (2023) 101712

Contents lists available at ScienceDirect

The Leadership Quarterly

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

Doing better leadership science via replications and registered reports

Melissa Carsten [a,*], Rachel Clapp-Smith [b], S. Alexander Haslam [c], Nicolas Bastardoz [d], Janaki Gooty [e], Shane Connelly [f], Seth Spain [g]

[a] College of Business Administration, Winthrop University, United States
[b] College of Business, Purdue University Northwest, United States
[c] School of Psychology, The University of Queensland, Australia
[d] Work and Organizational Studies, KU Leuven, Belgium
[e] University of North Carolina, Charlotte, United States
[f] University of Oklahoma, Norman, United States
[g] John Molson School of Business, Concordia University, Canada

ARTICLE INFO

Keywords:
Replication
Pre-registration
Leadership

ABSTRACT

In 2018, the *Leadership Quarterly* commissioned a Special Issue (SI) on replications of previously published studies in leadership. We adopted the Registered Reports path for this special issue to facilitate constructive reviewer feedback prior to data collection and to minimize any bias against non-significant findings in the publishing process. In this editorial, we reflect on how this approach worked, what we learned about replications, and identify three myths that we have seen throughout this process regarding the value and conduct of replications. We then offer actionable solutions to further advance replications in science. We close with a discussion of the six studies included in this SI and the implications of these studies for leadership science. We hope that this work continues to inspire replications and registered reports and in support *The Leadership Quarterly* offers an open call for such submissions.

Introduction

Replication is the cornerstone of the scientific method, yet replication studies are rare in the social and behavioral sciences (Block & Kuckertz, 2018; Köhler & Cortina, 2021; Makel, Plucker & Hegarty, 2012). This paucity of replication research has also been observed in many other fields of science — from psychology, economics, accounting and finance to biology as well as medicine — causing doubts regarding the credibility and relevance of major findings in such domains. Recently, calls have been made in the social, behavioral, and organizational sciences for greater emphasis to be placed on reproducible and replicable research (Gertler, Galiani & Romero, 2018; Munafo et al., 2017). This interest is at least partially driven by concerns about a replication crisis in science (Errington et al., 2021; Ioannidis, 2012; The Open Science Collaborative, 2015). The spirit of these recent efforts to reignite interest in replication research drove the inception of this special issue. Nosek and Errington (2020; p. 1) define replication as "repeating a study's procedure and observing whether the prior finding recurs" with new data. This is distinct from observing if a prior finding recurs via analytic reproducibility (same data, same analysis) and robustness checks (same data, different analytic assumptions; Hardwicke et al., 2018).

This special issue assembles a collection of studies which aim to provide *results-blind* replications of several key research findings in the field of leadership. This process was managed via Registered Reports – a system that allows authors to submit proposals, receive feedback from reviewers, and obtain an in-principle acceptance decision prior to data being collected (Chambers, 2019; Chambers & Tzavella, 2022; Grand et al., 2018). Registered reports guard against investigator dependencies in replication results (e.g., those pertaining to experimenter bias; Rosenthal, 1966; Simmons et al., 2011) as well as biases that arise from publishing only "interesting" or "significant" findings (Antonakis, 2017; Ozonoff, 2011; Scheel, Schijens & Lakens, 2021). Critically, a Registered Report process provides the opportunity for constructive review feedback prior to the beginning of data collection rather than after the fact. Although considered the "gold standard" in replication research (Chambers et al., 2014), the registered reports process was both time and labor intensive for all involved – the study authors, reviewers and the seven action editors expended more effort than is typical for producing a SI. This special issue is thus five years of a labor of love for all of us and many lessons learned in the process.

This special issue is a modest step in promoting replication research in the leadership field; yet as editors, the production process exposed

* Corresponding author.
E-mail address: carstenm@winthrop.edu (M. Carsten).

https://doi.org/10.1016/j.leaqua.2023.101712
Available online 7 June 20231048-9843/

many challenges to the continued conduct of replications of which we had not previously been aware. First, replication as a methodology is not well understood and requires substantial effort and application to do well. Indeed, most scholars in our field simply are not trained to design, conduct or review replications. This lack of training may create unease at the prospect of conducting a replication study (at best), or antipathy towards the practice of replication (at worst). Second, these problems are exacerbated by the fact that replication is largely devalued in leadership studies, and seen as insufficiently lofty to warrant publication (especially in a top-tier journal; Madden, Easley, & Dunn, 1995). Third, and most importantly, science advances when repeated replications occur and some theories are retired, some are pruned, and others are reinforced. Yet, as a field, we collectively value theory building over theory testing and replication, and our "collective mind" as leadership scholars drive those who are party to the research process away from replication rather than towards it[1].

Given these three observations, we first clarify the nature and content of replications, followed by a discussion of the collective mind that reinforces myths regarding replications, strategies to combat such myths and close with a discussion of the six studies and how they contribute to advancing leadership science.

What is replication and why do it?

Köhler and Cortina (2021) provided a useful frame of reference for classifying different forms of replication (see also Schmidt, 2009)—ranging from studies which involve direct or close replication in which the exact methods and analyses are performed by a new research team, to those that involve conceptual replication in which the same basic research question or hypotheses are studied using a different method or analysis strategy (Schmidt, 2009). Speaking to this point, Köhler and Cortina (2021) described a continuum that ranges from literal replication (direct or close replication) to constructive replication, which involves changes to aspects of the original research design. They further distinguish between three types of constructive replication which vary in scale and ambition: incremental (which addresses one or more minor flaws in the original study), substantial (which attempts to identify and correct some, but not all, of the most important weaknesses), and comprehensive (which attempts to resolve all important weaknesses of the original study). In their review, Köhler and Cortina (2021) concluded that substantial constructive replication, although underutilized, is likely to make the most important contributions toward advancing and refining management theory.

Constructive replication allows for an investigation into elements of the original research that are likely to have shifted over time, finding sources of variability that may or may not have endured, and targeting those sources of variability in the design of replication studies (Köhler & Cortina, 2021). For example, two studies in this special issue explore the possibility that in light of recent historical developments (e.g., the Black Lives Matter movement), the "white standard" specified within leadership categorization theory (Rosette, Leonardelli, & Phillips, 2008) may no longer hold (Obenauer & Kalsher, this issue; Ubaka et al., this issue). Similarly, two studies examine potential variability in previously reported findings by utilizing samples that have greater gender diversity than the original studies (Hammond et al., this issue; Hopp & Pruschak, this issue). In this way, the various efforts at replication seek to retain the virtues of original studies while consolidating and clarifying their conclusions amidst a changing societal context.

In the process of editing this special issue, our editorial team grappled with these differing nuances and types of replications, which ultimately led to our second major insight. As leadership scholars, we are constrained by the limits of our collective mind, and the myths we propagate regarding replication research and its place in leadership research. In the next section, we attempt to both spell out (and perhaps dispel) these myths.

The collective mind and the myths it propagates

Weick and Roberts (1993) defined collective mind as a way of thinking that is based on patterns of collective behavior rather than the knowledge or expertise of individuals. As they observed in their influential study of naval personnel on aircraft carriers, the performance of a collective is grounded in the heedful (i.e., purposeful, careful, consistent, and conscientious) interrelated activities of its members. These interrelated activities create social forces that are part of a social system, and motivate future actions aligned with that system. In this way, the actions of individuals both create, and are created by, the system of social forces that together represent the collective mind.

Aligned with the thinking of the broader research communities of which they are part, the collective mind of leadership and management scholars have developed a set of assumptions about replications that guides our collective behaviors and research priorities. Our collective mind as leadership researchers has accepted and reinforced these assumptions, deemphasizing its rightful place in leadership science. We pinpoint some of these implicit assumptions, which are arguably myths, below.

Myth #1: Replication is adversarial

The idea that replication is necessarily an adversarial process can be seen to emerge from the way that replication is defined and experienced in the physical sciences — where failure to achieve direct replication is commonly regarded as "disproving" the original findings (Ioannidis, 2012). In the social and behavioral sciences, however, there is limited evidence that authors engage in replication primarily to "disprove" or "invalidate" previous studies (Köhler & Cortina, 2021). Isager (2018) reviewed 85 studies in the Curated Replications Database and found that the top five reasons given for choosing to replicate a particular study included (1) the theoretical significance of the original findings, (2) personal interest in the findings and a passion to study them in greater depth, (3) the overall impact of the study on the academic field of interest (e.g., as evidenced by citation counts and the study's status as a foundation for subsequent research), (4) the impact of the original study on society and public policy, and (5) methodological questions associated with the original study (e.g., a desire to estimate effect sizes). We observed similar drivers for the replications attempted in our special issue in so far as authors' primary (stated) motivation was the overall impact of the original study on the leadership field. These observations suggest that motivations for replication have more to do with the impact of a given study for a field than they do with a desire to be adversarial.

Myth #2: Replication is not valuable

Authors of the culminating report of the Open Science Collaborative (2015) noted that replications are highly complex and dependent on context, sample, and methodology, and that this makes the task of reproducing original conditions, and original findings, extremely challenging. In addition, interpreting replication results could be more nuanced than simply examining whether $p < .05$ (e.g., are effect sizes within the 95% CI of original findings? Is the effect size reduced, and if so, how much? What are the differences in power analysis between the original and replicated study?). Given that replication studies are unlikely to yield exactly the same results

[1] Recognizing these challenges with replications and our obsession with novelty, at least one new journal - *Journal of Management Scientific Reports* was launched precisely for theory testing and replications, during the five years we worked on this special issue.

as the original research, one may reasonably ask whether they are necessary at all (Köhler & Cortina, 2021). Although it remains a responsibility of the replicating author team to convey why a specific replication is critical, we argue that replicating important work is necessary and valuable for multiple reasons and encourage such research being published in flagship journals.

As we indicated in the call for papers for our special issue (Clapp-Smith et al., 2018), there are at least three reasons why replications are necessary and valuable in the leadership (and management) field. First, as noted above, the world is changing and this means that phenomena and relationships may have evolved since a given finding was first reported in ways that we want to discover and understand. For instance, researchers may want to understand the implications of organizations having become flatter and having more flexible structures, and of having more women in formal leadership roles. Second, our statistical knowledge has evolved in ways that currently allow researchers to test theories in more robust and valid ways. Third, the prevalence of endogeneity in leadership research (Antonakis et al., 2010; 2016; 2021) suggests that we need to use alternative techniques to make causal claims capable of better informing policy (see Klonek, Gerpott, & Parker, this issue).

Nevertheless, due to the emphasis our field places on novelty, originality, and theoretical innovation (Hambrick, 2007; Ioannidis, 2012), many researchers continue to believe that replication is not valuable (e.g., see King, Goldfarb, & Simcoe, 2021). Speaking to this point, Neuliep and Crandall (1991, 1993) observed that a majority of journal editors (72%) and reviewers (54%) in leading social and behavioral science journals found replication work to be less valuable than tests of novel ideas — the implicit belief here being that replication work only reproduces knowledge that we already have. Yet, as Muthukrishna and Henrich (2019) noted, this belief reflects an impoverished and impoverishing view of the scientific process that devalues the cumulative testing of established ideas and theory and that encourages the proliferation of narrowly focused mini-theories (see also Chalmers, 2013; Haslam & McGarty, 2019). This devaluation then extends to how academic careers are built and rewarded.

Early in their academic careers, researchers are commonly told to carve out an identity for themselves by finding a niche in their field of interest and making contributions within a particular research space. Moreover, given that few receive rigorous training in replication, it is understandable that replication work is rarely part of that space and hence rarely part of budding academics' research identity. Instead, replication is commonly seen as "someone else's job." As Makel, Plucker, and Hegarty (2012) noted, "a field that replicates its work is rigorous and scientifically sound, but researchers who conduct those replications are looked down on as bricklayers and not advancing knowledge" (p. 537). Authors may thus resist replication work because it fails to advance their identity as an expert in a specific realm. Likewise, reviewers may feel that replications are not in their methodological wheelhouse, and therefore decline invitations to review the replication work of others. By the same token, journal editors may feel that replications could receive fewer citations or attention (van Witteloostuijn, 2016), as authors are prone to citing the original work.

To dispel this myth, we need to recognize that there is often a tendency to assume a few experts and scientific leaders drive most of the scientific progress in a field. Yet, this reasoning omits the critical role played by the broader scientific community and its distributed activities.

Encouraging, appreciating, and rewarding bricklayers — and everyone who contributes to the collective process of knowledge building — is one way in which we can start to shift our collective mind. More generally, we need to tell new "vivid stories" that reward and model new forms of collective identity (Weick & Roberts, 1993, p. 367). In particular, we see opportunities to become more intentional in wrapping replication into the norms of the research process in our field.

Several journals have already begun the process of normalizing and reinforcing replications. The *Academy of Management Discoveries* now explicitly states that replications are part of the journal's mission. A new journal, *Journal of Management Scientific Reports* (a sister journal of *Journal of Management*) was founded for the express purpose of *theory testing rather than theory creation*, especially through replication studies. In similar vein, editorials in *The Leadership Quarterly* also noted the importance of replication for ensuring not only the credibility of our research but also the overall health and stability of our field (Antonakis, 2017; 2023). Although it may have been common in the past to equate replication research with inconsequential scholarship, as Bob Dylan famously said, "the times they are a changin."

Myth #3: Replication studies are boring

Management research has been criticized for an over-reliance on significant findings in empirical research (Antonakis, 2017; Mathieu, 2016; O'Boyle, Banks & Gonzalez-Mulé, 2017). Editors and reviewers typically do not find value in publishing null findings, and the merits of theory testing are commonly perceived through the lens of statistical significance (Antonakis, 2017). In replication research, the value of significant vs. non-significant findings is often harder to judge. There are multiple reasons why an original study may not replicate: for example, the societal context may have changed (van Bavel et al., 2016), the sample may be different, or the analyses and tests may be more powerful. As mentioned earlier, there is also value in examining effect sizes in the original and replicated study, that go beyond our typical definitions of statistical significance. Thus, whereas a successful replication – reporting that findings are reproduced – can add credence to claims in the original research, unsuccessful replications are inherently difficult to interpret because they can neither refute nor negate the original findings and claims (Brandt et al., 2014; Nosek & Errington, 2020).

In this special issue, we attempted to avoid significance bias by relying on registered reports (see Chambers, 2019; Chambers & Tzavella, 2022). Further, registered reports give the opportunities for authors to receive feedback prior to data collection which should improve the quality of the research process over and above a standard report. In doing so, we divided the publication process into a pre- and a post-data collection stage. In the pre-data collection stage, authors submitted an introduction including a justification for replication, a design for the replication study, and an analysis plan. At that stage, the review team assessed the justification for the replication. More importantly, we offered suggestions for improvement and refinement of the study design and analysis plan (Mellor & Nosek, 2018). Once the editorial team was satisfied with the revised registered report, we accepted manuscripts in principle (i.e., an In-Principle Acceptance) with the only caveat being that authors should deliver on their promise and implement what they had agreed.

In the post-data collection stage, the review team only assessed whether there was fidelity to the proposed design and whether results and findings were appropriately reported and discussed. In this way, registered reports protect the replicating author team from significance bias because the publication of research is independent of the research findings. This ensures that science advances on the basis of rigorous theory-based prediction, design, testing, and analysis (Banks et al., 2018; Bakker et al., 2012, Franco, Malhotra & Simonovits, 2016; Hollenbeck & Wright, 2017; van't Veer & Giner-Sorolla, 2016). Reflecting on the systems aimed at increasing methodological rigor of published work (including registered reports), Vazire (2018) states, "this [change] would necessarily make the published literature less exciting (fewer bold, surprising claims), but I believe sacrificing this type of creativity for greater accuracy is worth it" (p.414). We agree wholeheartedly.

Actionable solutions to encourage more replications

Identifying the assumptions that drive our collective mind is only the first step towards creating new norms around replication research. Without actionable steps forward, there is little hope that we might begin shifting our collective behavior toward increased practice of replication. Thus, we identified three solutions that could help dispel the myths noted above and move our field forward via robust replications of existing studies. In what follows, we elaborate briefly on each of these in turn.

Solution #1: Align our behaviors as leadership scholars with the practices of open science

According to The Open Science Project (Gezelter, 2009), the goal of open science is to increase transparency and accessibility throughout the entire research cycle from idea generation, through design and data analysis, and eventual publication. One of the main benefits of open science is that it helps a given field to achieve more consistent and efficient replication of key research findings. For example, the open science model encourages researchers to make their data available through open-access repositories, whereby detailed accounts of research routines, measures, workflow management systems, or experimental processes and coding are made available for other researchers to review and reproduce. In allowing open access, not only do researchers improve the transparency of their methodology and data collection, but easy access to methods and scientific data also improves the likelihood that such data will be reused in replication or other studies to improve the validity of scientific knowledge.

Evidence indicates that these open science practices have a number of benefits including the rate of successful replications (Protzko et al., 2023). Journals should consider adopting the Transparency, Openness, and Promotion guidelines (Nosek et al., 2015). Thousands of journals across the sciences and all major publishers have become signatories of these standards. Most major journals in management and applied psychology are now indexed in the TOP standards which evaluates journals for their open science practices as a complement to the traditional journal impact factor (JIF). The JIF solely reflects the citations received by journals which is a type of popularity metric like television ratings (Kepes, Banks, & Keener, 2020). Currently *The Leadership Quarterly* is in the 99th percentile of all indexed journals across the sciences at the time of this writing which supports the notion that a journal can be successful in terms of both the JIF and the TOP metrics.

Solution #2: Recognize and engage with barriers to the alignment of our identities with the practices of open science

As we know from social and organization research, people generally seek to maintain a clear and consistent link between their past, present, and future, and this desire for identity continuity leads any threats of change to their valued collective identities to be perceived as an existential threat (Haslam et al., 2008; Sani et al., 2007). Among leadership scholars, the desire for social identity continuity can be seen as a powerful force that drives a reverence for established ways of conducting, reporting, and publishing research. In itself, identity continuity may not be a bad thing because it has repeatedly been shown to increase well-being whereas discontinuity increases stress and a nostalgic longing for 'the way things used to be' (Iyer & Jetten, 2011). Yet this same nostalgia can also impede the process of adaptive change, in part because it motivates group members to 'dig their heels in' and obstruct progress (Jetten & Hutchison, 2011, Jetten & Wohl, 2012, Wohl et al., 2012). It follows, then, that if we want to make progress — of a form that improves the validity, credibility and utility of research — we need to encourage and promote healthy forms of identity change.

Solution #3: Appreciate the challenges of alignment as partly challenges of leadership — That we as leadership scholars should study and own

The previous two strategies speak to the fact that the process through which leadership scholars come to engage more constructively and more enthusiastically with replication as a core part of their science is a process of change. More specifically, we can see that the process here is one of identity change, and that many of the barriers relate to the nature of pre-existing identities and the forces that hold them in place. One process that has a key role to play in helping people navigate such change is leadership. In particular, Muhlemann and colleagues (2022) points to the importance of *identity leadership* in helping people to see identity change not as a threat but as an opportunity for personal and collective growth in which their primary experience is not one of identity loss but instead one of identity gain.

As an editorial team, we are animated by the prospect of providing some of this leadership, we are also excited by the particular relevance of these issues for our field as a whole. Indeed, leadership itself has an important role to play in helping us to address and overcome the replication crisis (and a wealth of other similar challenges associated with problems of scientific inertia). This special issue, then, invites leadership scholars not only to get on board the bus of change but also to reflect on the ways that their skills and insights as leadership scholars should allow us to understand how that bus is being driven and find ways to do this better.

Overview of papers in the special issue

This special issue consists of six papers, collectively producing 18 studies that replicate research relating to five different theories. It is tempting here to summarize how many studies replicated the findings of the original studies, but if you have been taking heed of this editorial, you would eschew any such summary, knowing it to be too simplistic (see Van Bavel et al., 2016). As we noted above, the goal of replication is not to provide a tally of hits and misses but to add relevant data points to the pool of empirical knowledge and theory. What follows is a brief overview of the six papers in this special issue that contribute to this goal.

These replication studies demonstrate the continuum of replication approaches from direct (or close) replications, to constructive as outlined by Köhler and Cortina (2021). Three studies attempted direct replications (Hopp & Pruschak, 2020, Hammond et al., 2021 Study 1; Ubaka, Lu, & Gutierrez, 2022). Of these three, one was an official reproducibility study (rather than a replication) and utilized exactly the same data and sample as the original study (Hopp & Pruschak, 2020) and two studies collected data utilizing the same methods with samples similar to those in the original study (Hammond et al.; Ubaka et al., 2022). The remaining fifteen studies took the form of independent constructive replications. Among these, a number of studies provided evidence for the important role that replication plays in keeping scientific findings relevant over time. Specifically, researchers identified classic studies in leadership (many of which were over 15 years old), and tested whether original findings remained the same despite societal, economic, and demographic changes.

The six articles compiled in this special issue provide important empirical evidence to consolidate our understanding of the leadership process. Hopp and Pruschak, 2020 replicated a study by Kuhn and Weinberger (2005) that found high school leadership roles (as team captain and/or class president) are positively related with earnings later in life. The authors added greater context and understanding to the original findings by (a) using more robust methodological strategies that correct for potential endogeneity (e.g., using propensity score matching and instrumental variable estimation), and (b) expanding the sample to include females and non-white males. In doing so, they successfully replicated the original results with white men. This

replication study adds knowledge to theories on diversity in leadership, how early leadership experiences influence earnings later in life, and raise new questions about the role of early leadership experiences for majority vs. minority groups.

Klonek and colleagues conducted a substantive constructive replication that tested four hypotheses derived from the broad body of ambidextrous leadership theory (Rosing, Frese, & Bausch, 2011). This substantive constructive replication identified opportunities to refine and strengthen ambidextrous leadership theory (ALT) as a result of the fact that "previous research suffers from problems with causal interpretation and endogeneity concerns" (p. 1). Because previous research relied exclusively on questionnaires, the authors developed two different experiments (a vignette experiment and a video-based) to manipulate ambidextrous leadership behaviors. Using an instrumental variable estimation procedure allowing for causal claims, the authors found very limited evidence for the effectiveness of ALT compared to other leadership behaviors (i.e., opening or closing behaviors alone, transformational leadership behaviors) when predicting innovation outcomes. These null findings suggest three possibilities: (1) Endogeneity might have affected the original findings which provided strong support for ALT, (2) Over-reliance on scales to operationalize ALT and (3) deficiencies in the ALT concept itself (i.e., tautological, outcomes-based). The authors concluded by offering important suggestions for future research using ALT, arguing in particular that the key constructs of exploration and exploitation need to be more clearly specified and repeated testing needs to occur.

In their contribution to this special issue, Hopp, Wentzel, and Roles reported an incremental constructive replication of a study by Rule and Ambady (2008). The original study found a positive correlation between naïve observers' inferential ratings of a company CEO's face (e.g., ratings of their leadership ability, competence, and attractiveness — based on a picture of the CEO) – and firm performance. Hopp and colleagues departed from the original study with a view to being able to make causal claims. First, "the original focused on the CEOs from the 25 highest and 25 lowest ranked Fortune 1000 companies (Hopp et al., this issue, p. 4.), which may have artificially increased correlations. Second, the empirical evidence linking leader faces and outcomes mostly relied on cross-sectional designs, preventing defector causal claims. Thus, while closely replicating the methods and measures used in the original study, Hopp and colleagues extended the original findings by increasing the sampling frame and the time horizon of evaluation. Even though they were able to replicate the original findings while using the original modeling strategy, they were not able to replicate findings when using more sophisticated econometric specifications (e.g., using firm fixed-effects, a different measure of the dependent variable). Their findings add to the existing literature by suggesting that while CEO face and observer ratings might covary, we cannot infer that CEOs facial appearance (and observer inferences) *cause* firm performance.

Interestingly, we accepted two proposals to replicate original research reported by Rosette, Leonardelli and Phillips (2008) who found evidence of a "white standard" in perceptions of leadership. Ubaka, Lu, and Gutierrez argued that socio-historical events (notably, the election of Barack Obama as the first black president of the U.S.) might have contributed to a change in implicit leadership theories, and in particular, those relating to race. To explore this idea, they used an incremental constructive approach with a view to advance and refine leader categorization theory (Lord et al., 2020). Although the original findings were only partially supported, the authors did find support for core elements of the underlying theory — notably the idea that whereas leader categorization is still a relevant process, the content of categorization may have changed.

Obenauer and Kalsher also replicated the research of Rosette, Leonardelli, and Phillips (2008). The authors conducted six studies to address limitations of the original experimental design (e.g., including better manipulation checks), and including non-student samples in tests of their hypotheses. Rather than finding evidence of a "white standard" in the categorization of leaders, the authors found evidence of the very opposite effect –namely a preference for non-white leaders. The authors argue that sensitivity to changes in social context, and the current significance placed on avoiding racial bias, may have influenced participant responses. Although the findings were fairly consistent across all six studies, the authors note that leadership categorization theory, and the tendency to see white males as leaders, may not have changed much over time (see also Petsko & Rosette, 2023). If the theory itself was no longer viable, the authors noted that they would have seen a reduction in participants' categorization of white males into leadership positions. Instead, they found that non-white individuals were categorized into leadership positions more frequently than in the original study. These two replications of the original study, combined, with Petsko and Rosette (2023) demonstrate the power of replications for our field – taken together, they suggest an extension of Leader Categorization Theory's core position regarding the white male standard. This white male standard has not faded but might be harder to detect via direct surveys due to social desirability concerns.

Finally, Hammond and colleagues conducted four studies to provide both direct and incremental constructive replications of the first study in Meindl and Ehrlich's (1987) research on the romance of leadership. The authors expand on the original study by broadening the sample in terms of age and work experience, and adding a number of experimental conditions accounting for the gender of the leader, and both organizational success and failure. The findings suggest that the tendency to engage in attribution bias by holding leaders primarily responsible for organizational success and failure has likely diminished in the past 30 years. This is important not only for the Romance of Leadership theory, but for leadership attribution theory more generally. The authors noted that their findings can be seen as reflecting progress in the way society has come to define leadership and its importance to organizational outcomes.

Conclusion

The process of editing this special issue made it clear to us that replication is a complex research process that requires scrupulous intentional and heedful attention, rigorous training, and meticulous planning. In the process we also identified three key myths that have stymied replication research and thereby hampered progress in the study of leadership. Ultimately, this raises the question of whether classical leadership theories endure the test of time, and whether such theories operate today the same way they were theorized many years ago.

As scientists, it is our responsibility to ask probing questions, but not all such questions are novel and edgy. Here, then, there needs to be space to re-examine the foundations of our science and to shift our focus from what is bright and new onto what has been taken (often incorrectly) as familiar truth (Smith & Haslam, 2017; Steffens et al., 2022). Certainly, if the state of our science were measured in terms of empirical and theoretical innovation, it would be accounted overwhelmingly strong. However, if it were measured by theoretical development, rigorous testing, retesting, and refinement (as is the case in many other sciences), then we have some work to do (Muthukrishna, & Henrich, 2019).

The replication studies in this special issue make a compelling case for the value of this work. Each was conducted with a high level of rigor, meeting a high threshold for study design as a result of the scrutiny to which the proposed research was exposed during the registered review process. Moreover, the findings of each study provide nuanced insights into important theories and studies in the leadership field. In the process, they serve as useful models for leveraging replication as a critical research methodology. In light of this, we have suggested three

strategies that our field should look to implement in order to advance our science. With these various objectives and strategies mapped firmly in our collective mind, we hope that this special issue serves as a launching pad for the engagement of our community in concerted dialog about how we can all help to normalize replication in our field so as to ensure that we valorize the solid and the true not just the shiny and the new. We hope that this work continues to inspire replications and registered reports and in support *The Leadership Quarterly* offers an open call for such submissions.

References

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

Antonakis, J. (2023). In support of slow science: Robust, open, and multidisciplinary. *The Leadership Quarterly*. https://doi.org/10.1016/j.leaqua.2023.101676, 101676.

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., 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., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods, 24*(2), 443–483.

Bakker, M., Van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*(6), 543–554.

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.

Block, J., & Kuckertz, A. (2018). Seven principles of effective replication studies: Strengthening the evidence base of management research. *Management Review Quarterly, 68*(4), 355–359.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology, 50*, 217–224.

Chalmers, A. F. (2013). *What is this thing called science?* (4th ed). Hackett Publishing.

Chambers, C. (2019). What's next for registered reports? *Nature, 573*(7773), 187–189.

Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of" playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience, 1*(1), 4–17.

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of registered reports. *Nature Human Behaviour, 6*(1), 29–42.

Clapp-Smith, R., Carsten, M., Gooty, J., Connelly, S., Haslam, A., Bastardoz, N., et al. (2018). Special registered report issue on replication and rigorous retesting of leadership models. *The Leadership Quarterly, 29*(2).

Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *eLife, 10*.

Franco, A., Malhotra, N., & Simonovits, G. (2016). Underreporting in psychology experiments: Evidence from a study registry. *Social Psychological and Personality Science, 7*(1), 8–12.

Gertler, P., Galiani, S., & Romero, M. (2018). How to make replication the norm. *Nature, 554*(7693), 417–419.

Gezelter, D. (2009). What, exactly, is open science?". The Open Science Project, available at: https://openscience.org/what-exactly-is-open-science/.

Grand, J. A., Rogelberg, S. G., Banks, G. C., Landis, R. S., & Tonidandel, S. (2018). From outcome to process focus: Fostering a more robust psychological science through registered reports and results-blind reviewing. *Perspectives on Psychological Science, 13*(4), 448–456.

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal, 50*(6), 1346–1352.

Hardwicke, T. E., Mathur, M., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., & Tessler, M. H. (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*.

Haslam, C., Holme, A., Haslam, S. A., Iyer, A., Jetten, J., & Williams, W. H. (2008). Maintaining group memberships: Social identity continuity predicts well-being after stroke. *Neuropsychological Rehabilitation, 18*(5–6), 671–691.

Haslam, S. A., & McGarty, C. (2019). *Research methods and statistics in psychology* (3rd ed.). Sage.

Hammond, M. M., Schyns, B., Lester, G. V., Clapp-Smith, R., & Thomas, J. S. (2021). The romance of leadership: Rekindling the fire through replication of Meindl and Ehrlich. *The Leadership Quarterly*.

Hollenbeck, J. R., & Wright, P. M. (2017). Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management, 43*(1), 5–18.

Hopp, C., & Pruschak, G. (2020). Is there such a thing as leadership skill?–A replication and extension of the relationship between high school leadership positions and later-life earnings. *The Leadership Quarterly*.

Hopp, C., Wentzel, D., & Rose, S. (2020). Chief executive officers' appearance predicts company performance, or does it? A replication study and extension focusing on CEO successions. *The Leadership Quarterly*.

Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7*(6), 645–654.

Isager, P.M. (2018). What to replicate? Retrieved from https://pedermisager.org/blog/what-to-replicate/ on October 15, 2022.

Iyer, A., & Jetten, J. (2011). What's left behind: Identity continuity moderates the effect of nostalgia on well-being and life choices. *Journal of Personality and Social Psychology, 101*(1), 94–108.

Jetten, J., & Hutchison, P. (2011). When groups have a lot to lose: Historical continuity enhances resistance to a merger. *European Journal of Social Psychology, 41*(3), 335–343.

Jetten, J., & Wohl, M. J. (2012). The past as a determinant of the present: Historical continuity, collective angst, and opposition to immigration. *European Journal of Social Psychology, 42*(4), 442–450.

Kepes, S., Banks, G. C., & Keener, S. (2020). The TOP factor: An indicator of journal quality to complement Journal Impact Factor. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 13*, 328–333.

King, A., Goldfarb, B., & Simcoe, T. (2021). Learning from testimony on quantitative research in management. *Academy of Management Review, 46*(3), 465–488.

Klonek, F. E., Gerpott, F. H., & Parker, S. K. (2020). A conceptual replication of ambidextrous leadership theory: An experimental approach. *The Leadership Quarterly*.

Köhler, T., & Cortina, J. M. (2021). Play it again, Sam! An analysis of constructive replication in the organizational sciences. *Journal of Management, 47*(2), 488–518.

Kuhn, P., & Weinberger, C. (2005). Leadership skills and wages. *Journal of Labor Economics, 23*(3), 395–436.

Lord, R. G., Epitropaki, O., Foti, R. J., & Hansbrough, T. K. (2020). Implicit leadership theories, implicit followership theories, and dynamic processing of leadership information. *Annual Review of Organizational Psychology and Organizational Behavior, 7*, 49–74.

Madden, C. S., Easley, R. W., & Dunn, M. G. (1995). How journal editors view replication research. *Journal of Advertising, 24*(4), 77–87.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*(6), 537–542.

Mathieu, J. E. (2016). The problem with [in] management theory. *Journal of Organizational Behavior, 37*(8), 1132–1141.

Meindl, J. R., & Ehrlich, S. B. (1987). The romance of leadership and the evaluation of organizational performance. *Academy of Management journal, 30*(1), 91–109.

Mellor, D. T., & Nosek, B. A. (2018). Easy preregistration will benefit any research. *Nature Human Behaviour, 2*(2). 98–98.

Mühlemann, N., Steffens, N. K., Ullrich, J., Haslam, S. A., & Jonas, K. (2022). Understanding responses to an organizational takeover: Introducing the Social Identity Model of Organizational Change. *Journal of Personality and Social Psychology, 123*(5), 1004–1023.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour, 3*(3), 221–229.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), 1–9.

Neuliep, J. W., & Crandall, R. (1991). Editorial bias against replication research. *Journal of Social Behavior and Personality, 5*(4) 85.

Neuliep, J. W., & Crandall, R. (1993). Reviewer bias against replication research. *Journal of Social Behavior and Personality, 8*(6) 21.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., et al. (2015). Promoting an open research culture. *Science, 348*, 1422–1425.

Nosek, B. A., & Errington, T. M. (2020). The best time to argue about what a replication means? Before you do it. *Nature, 583*(7817), 518–520.

Obenauer, W. G., & Kalsher, M. J. (2022). Is white always the standard? Using replication to revisit and extend what we know about the leadership prototype. *The Leadership Quarterly*.

O'Boyle, E. H., Jr, Banks, G. C., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management, 43*(2), 376–399.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716.

Ozonoff, S. (2011). The first cut is the deepest: Why do the reported effects of treatments decline over trials? *Journal of Child Psychology and Psychiatry, 52*(7), 729–730.

Petsko, C. D., & Rosette, A. S. (2023). Are leaders still presumed white by default? Racial bias in leader categorization revisited. *Journal of Applied Psychology, 108*(2), 330–340.

Protzko, J., Krosnick, J., Nelson, L. D., Nosek, B. A., Axt, J., Berent, M., Buttrick, N., DeBell, M., Ebersole, C. R., & Lundmark, S. (2023). *High replicability of newly-discovered social-behavioral findings is achievable. psyarxiv*.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton Century Crofts.

Rosette, A. S., Leonardelli, G. J., & Phillips, K. W. (2008). The White standard: Racial bias in leader categorization. *Journal of Applied Psychology, 94*(4), 758–777.

Rosing, K., Frese, M., & Bausch, A. (2011). Explaining the heterogeneity of the leadership-innovation relationship: Ambidextrous leadership. *The Leadership Quarterly, 22*(5), 956–974.

Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science, 19*(2), 109–111.

Sani, F., Bowe, M., Herrera, M., Manna, C., Cossa, T., Miao, X., & Zhou, Y. (2007). Perceived collective continuity: Seeing groups as entities that move through time. *European Journal of Social Psychology, 37*(6), 1118–1134.

Scheel, A. M., Schijen, M. R., & Lakens, D. (2021). An excess of positive results: Comparing the standard Psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science, 4*(2). 25152459211007467.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*(2), 90–100.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366.
Smith, J. R., & Haslam, S. A. (2017). *Social psychology: Revisiting the classic studies* (2nd ed.). Sage.
Steffens, N. K., Ryan, M. K., & Rink, F. (2022). *Organizational psychology: Revisiting the classic studies*. Sage.
Ubaka, A., Lu, X., & Gutierrez, L. (2022). Testing the generalizability of the white leadership standard in the post-Obama era. *The Leadership Quarterly*.
Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences, 113*(23), 6454–6459.
Van Witteloostuijn, A. (2016). What happened to Popperian falsification? Publishing neutral and negative findings: Moving away from biased publication practices. *Cross Cultural & Strategic Management, 23*(3), 481–508.
Van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of experimental social psychology, 67*, 2–12.
Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science, 13*(4), 411–417.
Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, 357–381.
Wohl, M. J., Squires, E. C., & Caouette, J. (2012). We were, we are, will we be? The social psychology of collective angst. *Social and Personality Psychology Compass, 6*(5), 379–391.