EXPLORING EMERGENT AND POORLY UNDERSTOOD PHENOMENA IN THE STRANGEST OF PLACES: THE FOOTPRINT OF DISCOVERY IN REPLICATIONS, META-ANALYSES, AND NULL FINDINGS

Late at night in dimly lit bars around the globe, many members of our field secretly pine for a world in which replications, meta-analyses, and null findings are valued as key ingredients in our field's quest for knowledge generation. As these individuals know all too well, such knowledge vehicles are often looked down upon as the playgrounds of the less gifted, the less creative. Replications and null findings in particular have been treated very poorly in our field, and many other fields as well. Meta-analyses have fared better, but they do not yield many best-paper awards or strong prestige.

Recently, however, a crisis of faith has swept across the general academic community (Bettis et al., 2016a; Ioannidis, 2005; Open Science Collaboration, 2015). Many people have woken up and discovered the emperor has no clothes, or at least not very many. Reproducability of findings has been the focus of the crisis, which has brought replications and meta-analyses a bit more to the foreground as possible saviors. The crisis has also elevated the status of null findings, as people have begun to see more clearly our need to know what does not have effects rather than just what does have effects. Of course, we also have begun to better appreciate the fact that a dichotomous focus on statistical significance carries nontrivial dysfunction (Bettis, 2012; McKee & Miller, 2015), which has very important implications for the treatment of null findings. This new-found appreciation for avoiding simple yes–no distinctions also has implications for the definition of success in a replication study. Moreover, it reinforces the longstanding focus of meta-analysis on effect sizes rather than statistical significance.

The Academy of Management launched *Academy of Management Discoveries* (*AMD*) with the intention that it would play a leading role in elevating the status of replications, meta-analyses, and null findings in the field of management. We view all three of these knowledge vehicles as fitting within the journal's broader mission, namely to disseminate findings regarding emerging and poorly understood phenomena that have important implications for downstream theory building and managerial practice. In this article, we explore the types of replications and meta-analyses that fit the *AMD* mission and highlight the important role of null findings as well.

A PRIMER ON *AMD* FUNDAMENTALS

As indicated previously, the mission of *AMD* is to disseminate research on emerging and poorly understood phenomena using a pretheory orientation, or in other words grounded in an abductive logic (Van de Ven, 2016). Emerging phenomena are those that have escaped our field's notice in the past or are new in the organizations that we study. Poorly understood phenomena are those that our field has failed to comprehend despite a number of attempts. A pretheory orientation suggests approaches to quantitative work (our focus here) that are based on hunches, observation, and simple logic, rather than elegant theoretical treatises, where such treatises might be based on deductions from existing grand theory (e.g., agency theory, prospect theory) and/or sophisticated deductive logic informed by such theory (Bamberger & Ang, 2015). A pretheory orientation is most appropriate when existing theory is not relevant or not easily applied. These are precisely the contexts in which a more abductive approach to theorizing is valid, where the focus is on empirical findings and what may be plausible rather than on a priori expectations and what is valid. Plausibility rather than validity is an *AMD* keystone.

These fundamental aspects of *AMD* have important and direct implications for the types of replications, meta-analyses, and null findings that are appropriate for the journal. With an eye toward creating a match between the particulars of *AMD* and the work submitted to the journal, we explore these knowledge vehicles in more detail next.

REPLICATION RESEARCH IN *AMD*

It is well established that reproducibility is at the heart of the scientific enterprise and critical to the development of any scientific field (Open Science Collaboration, 2015). However, because premier journals place a premium on novelty, innovation, and interestingness, research focused exclusively on replication is notoriously difficult to publish. The result is a "perfect storm" for problematic science. The combination of an implicit incentive to identify and then theoretically "predict" counterintuitive relationships with an implicit disincentive to replicate such findings has resulted in the emergence of

scientific literatures in which many findings may be open to suspicion (Open Science Collaboration, 2015). For example, replicating 100 studies in psychology, the Open Science Collaboration (2015) reported mean replication effect sizes at half the magnitude of those reported in the original studies, with just 36 percent of replicated relationship having statistically significant ($p < .05$) results (as opposed to 97 percent of the originally studied relationships). While these particular findings have been criticized, they are part of a larger pattern that has generated substantial concern (Bettis et al., 2016b).

Because replication studies are important in developing any research-based field, such studies are welcome at *AMD*. But what does this really mean? Will *AMD* be open to publishing any replication study, and if not, precisely what kind of replication does *AMD* seek to publish? Replication research submitted to AMD will be evaluated on the basis of three main sets of criteria (the three Ms) having to do with motivation, method, and meaning.

Motivation-Related Criteria

Consistent with the mission of *AMD* noted earlier, replications must be focused on emerging phenomena or poorly understood phenomena. Understanding is a matter of interpreting what is known. Most might believe a given phenomenon is well understood when it really is not, while in other cases most might agree that a phenomenon is not well understood.

Replications are particularly useful for understanding important emerging phenomena. Consider the following scenario. A study is conducted in a novel area, and a moderately large, statistically significant effect size is found for the focal relationship. The next study of the relationship would be the first attempted replication and would provide 50 percent of all available evidence (one of two studies), making it a quite powerful addition. The third study would be only the second attempted replication and would provide 33 percent of all available evidence (one of three studies), which also would add substantial value in the still novel territory. The 20th study, however, would provide only 5 percent of the available evidence (1 of 20 studies). The 20th study would be very important in our evidence-based field where cumulative knowledge is important, but that study would not be appropriate for *AMD* if it were simply another successful replication, or another in a series of studies that had consistently failed to reproduce the original strong effect size.

Replications are also very useful for important phenomena that appear to be well understood but for which hunches, observations, and simple logic suggest blind spots or aspects that are not truly well understood. For example, a negligible effect size (and therefore null findings) might be suspected for a particular context whereas moderately strong effect sizes have been found in previous studies. This suspicion might be focused on issues regarding the internal or external validity of an empirically established relationship that in fact has served as the basis for broad theorizing in our field or that has had company-wide or public policy implications. Employee drug testing is interesting in this regard. It has been widely justified on the basis of empirical research establishing the relationship between workforce substance use and workplace accidents and injuries, but Frone (2013) has suggested several reasons why such findings should be subject to more rigorous replication (e.g., the possibility that general use is not actually important in comparison to actual impairment at work). Another area that is attractive for replications relates to suspected low average levels of some important variable despite high levels of this variable having been claimed in a number of previous studies. Infurna and Luthar (2016: 175–176) noted that decades of research have suggested widespread resilience among people that has led to conventional wisdom that says "in the aftermath of events such as 9/11 or natural disasters, widespread prophylactic interventions are not just unnecessary but even harmful." They, however, had a hunch that prior findings were methodological artifacts. Relaxing two methodological assumptions, the findings of Infurna and Luthar (2016) indicated a far lower prevalence of psychological resilience. Their work has substantial implications for theories of psychological resilience and for related theories, as well as for public policy.

In general terms, authors should motivate their replication research not only by explaining the importance of an established relationship but also by clearly specifying why there is reason to question or reassess prior findings. Possible reasons might include concerns and hunches about methodological artifacts, flaws in research design, and alternative explanations. In addition, concerns about possible boundary conditions could provide the motivation for assessing the external validity of prior findings. It should be noted, however, that findings indicating unexpected reproducibility are no less important than more typical findings indicating irreproducibility. For example, while one might question the generalizability of findings regarding the adverse impact of rudeness on individual performance in normative contexts in which rudeness is more socially acceptable, replication research indicating reproducibility—that the effects are robust regardless of normative context (Riskin et al., 2015)—carry as much importance as findings indicating irreproducibility (i.e., null findings).

AMD is unlikely to publish replication research focused on a relationship that has already been subject to meta-analysis. Many established relationships in the social sciences have been the focus of dedicated replications (against all odds and perhaps in lesser journals) and also have been subject to indirect replication when the focal variables of earlier research have been included as covariates/controls in subsequent research. In this sense, meta-analytic research serves as a critical tool for leveraging and supporting replication research, examining the external validity of the established relationship and its suggested effect size and potentially even identifying boundary conditions and empirically specifying how the effect is conditioned by such factors. Accordingly, if a relationship has already been meta-analyzed on the basis of a reasonable number of studies covering a range of theoretically grounded contexts and boundary conditions, it would be difficult to establish the emergent nature or even poorly understood nature of the phenomenon.

Methodological Criteria

AMD's mission statement places a premium on state-of-the-art methodological rigor, with the primacy of rigor being generalizable to replication research as well. Demonstrating rigor is likely to be easier in the case of *direct replication* ("the attempt to recreate the [original] conditions"—*Science*) relative to *close replication* ("recreate a study as closely as possible"—Brandt et al. in *Journal of Experimental Social Psychology*). This is because in a direct replication, the data are likely to be identical or nearly identical to those originally analyzed (e.g., same publically available dataset), with the only modifications being, for example, some adjustment in the operationalization of a variable or mode of analysis. Ganzach (2016) provides an illustration in his replication research focused on cognitive ability and party identity in the United States. Using the same database as the original study, Ganzach demonstrates how the previous finding (i.e., that those with higher cognitive ability have a higher probability of identifying with the Republican Party) is no longer supported when theoretically grounded covariates are taken into account. However, even in this study, Ganzach reinforced his finding by including a close replication of his own work, demonstrating the same noneffect in a separate database using stronger measures.

Because reproducibility in close replications can be strongly influenced by a variety of methodological factors including the nature and size of the sample, the empirical context, the operationalization of the measures, and the mode of analysis (Gilbert, King,

Pettigrew, & Wilson, 2016), authors must make a compelling case for the validity of the replication. One of the best ways to make this case is to demonstrate the consistency of the findings across at least two separate datasets, preferably using multiple methods. Additionally, scholars should strive to follow the "ingredients" specified in Brandt et al.'s (2014: 218) "Replication Recipe," namely:

1. Carefully defining the effects and methods that the researcher intends to replicate.
2. Following as exactly as possible the methods of the original study (including participant recruitment, instructions, stimuli, measures, procedures, and analyses).
3. Engineering high statistical power.
4. Making complete details about the replication available, so that interested experts can fully evaluate the replication attempt (or attempt another replication themselves).
5. Engaging in careful evaluation of replication results, while comparing them critically to those of the original study.

In striving to use these ingredients, authors should offer a compelling explanation for decisions regarding sample size and statistical power, exclusion criteria and policies for handling outliers, and measures (and where relevant, procedures, manipulations, and analytical methods). Furthermore, authors should emphasize effect sizes and confidence intervals, and indicate how these confidence intervals overlap (or fail to do so) with prior findings [see, e.g., Starbuck (2006)].

Meaning-Related Criteria

Meaning-related criteria have to do with the implications of the replication results for theory and practice. As a journal dedicated to pretheory, a critical criterion for replication research in *AMD* concerns the potential impact of the findings for downstream theory development. Accordingly, authors should attempt to (1) highlight how their findings might influence theory directions going forward or theoretical assumptions, and (2) identify any new criteria suggested by their findings for future theory development. For example, from the evidence presented in their replication study of employee and customer perceptions of service in banks, Schneider and Bowen (1985) argued for theories of consumer perceptions and behavior in service contexts to more comprehensively take into account the influence of the attitudes and perceptions of those serving them, as well as for theories of service employee attitudes to pay closer attention to the impact of consumer affect

and behavior. Schneider and Bowen's leveraging of replication findings to direct downstream theory building and to other targets for future theorizing epitomizes the abductive logic to which *AMD* is dedicated.

META-ANALYTIC RESEARCH IN *AMD*

Meta-analytic research has recently evolved in some segments of our field from a technique dedicated to cumulating past research findings and examining boundary conditions to a technique for testing hypotheses deduced from existing grand theories [compare, e.g., Junni, Sarala, Taras, & Tarba (2013), from *Academy of Management Perspectives*, with D'Innocenzo, Luciano, Mathieu, Maynard, and Chen (2016), from *Academy of Management Journal*]. As such, for some, meta-analysis has transitioned from a tool critical for evidence-based management to a tool for theory development and testing. Which approach is most relevant for *AMD*? Neither! As explained in the following sections, we are positioned between these two extremes.

Motivation-Related Criteria

Consistent with the mission of *AMD*, meta-analyses must be focused on important emerging phenomena or already studied but poorly understood phenomena. While it may seem odd to suggest that meta-analyses could be used to study emerging phenomena, it actually is not.

Meta-analyses could be useful for the study of important emerging phenomena. Consider the following. A particular relationship has been studied a number of times, with a few direct replications and several close replications in the mix. The relationship also is present in other studies because the key variables have been used as covariates/controls. A previous meta-analysis might even exist. Although a great deal is known about the focal relationship, the past findings could be used in the exploration of a new phenomenon that is captured by a potential moderator for which previous empirical work does not exist *and* for which strong theory does not come into play. It is this nexus of no empirical work and lack of strong theory that would make the analysis relevant for *AMD*. For example, a set of authors might have a hunch that hormones play a key role in moderating the negative relationship between CEO tenure and firm performance [for an overview of existing tenure–performance research, see Meschi, Metais, and Miller (2015)]. Noting that testosterone and cortisol levels tend to be higher among CEOs in certain industries, higher among CEOs in certain parts of the world, and higher among CEOs in certain

periods, they might propose that effects of CEO tenure are less negative in some of these contexts. The authors then could use past studies of CEO tenure and performance to investigate their hunch using meta-analysis, where industries, regions, and time would be moderating factors. Given the dearth of research on hormonal effects in upper-echelon research but the latent interest and potential impact of such work (e.g., Apicella, Carre, & Dreber, 2015; Carney & Mason, 2010), the authors could offer an important contribution that might bring form and direction to future work in the upper-echelon tradition as well as a host of other management research streams.

Meta-analyses also can be very useful for long-standing but poorly understood research areas. Here, substantial variation in past findings, frustration, and perhaps partial abandonment of a given research area might be involved. A sense of defeat often is seen, as was the case in planning–performance research in the late 1980s and early 1990s. Mintzberg (1994) had gone so far as to conclude that planning research had been absolutely worthless, and he did so in an award-winning book. Meta-analytic work, however, showed that Mintzberg's characterization was not quite on target (see Miller & Cardinal, 1994). For *AMD*, meta-analyses in areas such as planning–performance in the 1990s would be particularly appropriate, assuming hunches, observations, and simple logic-driven investigations of methodological and substantive moderators in the absence of strong theory are used to guide the work.

Importantly, meta-analyses are particularly well-suited for examining suspected negligible effects by moderating variables. Because the underlying aggregate sample sizes tend to be quite large, statistical power tends to be very strong. Thus, the chances of type II errors are low. Any null findings are likely to be meaningful, even more meaningful than those reported in a very strongly powered primary study (which *AMD* also welcomes).

Methodological Criteria

As discussed previously, *AMD*'s mission puts a premium on state-of-the-art methodological rigor, and that certainly applies to meta-analytic research. The fundamentals of rigorous meta-analysis are well known (see Bornstein, Hedges, Higgins, & Rothstein, 2013; Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Schmidt & Hunter, 2015), but transparency and reproducability in applying those methods are key. According to the American Psychological Association's *Publication Manual* (see the JARS-MARS Appendix), such transparency includes, but is not limited to, disclosure of (1) the search methods used

in identifying samples, (2) qualifications of those who coded data, (3) noteworthy judgment calls in the coding and robustness tests used to gauge the impact of those calls, (4) use or nonuse of attenuation corrections, (5) any adjustments for outliers, and (6) an indication of the variance in a set of correlations attributable to sampling error. To achieve full reproducability, authors should consider sharing their data on the *AMD* website.

Beyond the fundamentals, authors must be clear in specifying how they handled several key issues and why they handled them in those ways. First, there is the question of fixed versus random effects. Analyzing these different types of effects serves very different purposes. For fixed effects, existing samples are considered to be the entire population of interest rather than being seen as an accurate representation of all possible samples that theoretically could have been created over time. Thus, conclusions can be drawn only for the samples at hand, which limits one's ability to view the parameter estimates as representations of true relationships in any broad sense (Borenstein, Hedges, Higgins, & Rothstein, 2009). For random effects, existing samples are assumed to have been randomly drawn from the universe of all possible samples that theoretically could have been created over time. If this assumption is a reasonable representation of reality, then random-effects results represent truth in a broad sense. In summarizing the plusses and minuses of the two approaches, Borenstein et al. (2013: 16–17) said this: "The fixed-effects model has good precision but is shooting at the wrong target (limited to the studies at hand) . . . . The random-effects model is shooting at the right target (the universe of studies) but with poor precision." Given this empirical landscape, authors must carefully explain their approach.

Second, there is the question of publication bias. Publication bias exists when published evidence is not representative of all existing evidence (Schmidt & Hunter, 2015). It can be generated by reviewer/editor bias in what to publish and/or by author decisions regarding what to submit to journals. In both cases, decisions might turn on statistical significance and/or empirical support for a particular ideology, with the former being the chief source of concern. As has been observed on many occasions, the rate of statistical significance seems quite high in our field given our emphasis on novel hypotheses and effects, and also given the questionable statistical power in many of our research streams (Schmidt & Oh, 2016). If bias exists, and certainly this is not always the case, then parameter estimates from the most elegantly executed meta-analyses likewise will be biased. To combat this, authors should consider several tactics including (1) comparing average effect sizes from replications to average effect sizes from tangential studies where variables of interest were included as covariates/controls and (2) comparing average effect sizes from published studies with average effect sizes from unpublished studies. Ideally, the average effect sizes in these comparisons would be similar. Authors also should conduct a so-called file drawer test to see how many additional studies with negligible effect sizes would bring the population estimate calculated in the meta-analysis into the negligible range. Ideally, the number would be quite large.

Third, there is the question of using unpublished research in the main analyses. Unpublished research has not been vetted (or successfully vetted) in the peer review system, and it is generally not available for use in theory building for primary studies, textbook writing, or classroom discussions. Put another way, it is not an official part of the actual stock of knowledge on which the field operates. Yet, using such research has become quite popular in meta-analytic studies. However, prior to using such research outside of a secondary check for publication bias, authors should carefully explain how/why the process of acquiring unpublished research yielded a representative sample of such research. In addition, they should control for the status of samples in their analyses (published versus unpublished), and/or run a sensitivity test to assess the impact that including such studies has on the overall effect size.

Meaning-Related Criteria

Meaning-related criteria have to do with the implications of meta-analytic findings. As is the case for any type of work submitted for possible publication in *AMD*, the potential impact of a meta-analysis on downstream theory building and managerial practice is a key consideration. Importantly, one of the benefits of meta-analytic work relates to the estimates of aggregate effect sizes based on thousands and thousands of observations, including effects for various contexts (e.g., small versus large organizations). Such effect sizes provide much more information than the dichotomous yes–no emphases of statistical significance. And these effect sizes can be used to calculate estimates of practical benefit. Using the binomial effect-size display (Rosenthal & Rubin, 1982), Lipsey and Wilson (1993) showed that an aggregate correlation of only 0.24 (where many of the individual effect sizes represented null findings due to low statistical power) meant a success rate of 38 percent for those low on the independent variable and a success rate of 62 percent for those high on the independent variable. That is a very robust and substantively important 63 percent increase.

Because of their potential to bring powerful and durable insights, meta-analyses often have substantial consequences for both theory and practice. As Eden

(2002: 844) pointed out a number of years ago, ". . . the findings of meta-analysis can raise new theoretical questions and frontiers . . . meta-analysis is not necessarily the terminus in a stream of research; it can also point to the best direction for new theory development and consequently for further replication research." To turn the potential for impact into actual impact, authors should highlight how their meta-analytic findings could be used in several research streams or theory traditions. They must speculate widely on the potential uses of their work in downstream theory building. This is what we mean at *AMD* by the phrase "implications for theory."

NULL FINDINGS

As mentioned earlier, null results are highly relevant for replications and meta-analyses. Null results, however, also should be considered important in their own right. Our field's preoccupation with statistically significant effects has produced a number of distortions related to publication bias and general neglect of important noneffects (Schmidt & Oh, 2016). Imagine if medical researchers were unconcerned with understanding treatments that do not work! At *AMD*, null findings are welcome, so long as they have been produced in a pretheory setting and with empirical methods that ensure sound measurement and strong statistical power. Purposeful searches for negligible effects also might benefit from Bayesian approaches (McKee & Miller, 2015).

CONCLUSION

Replications, meta-analyses, and null findings provide important ways to explore emergent and poorly understood phenomena. Their potential for advancing knowledge of management and organizations is strong. At *AMD*, we are committed to embracing these knowledge vehicles as we continue to help the field realize its potential for a bright future.

C. Chet Miller
University of Houston

Peter Bamberger
Tel Aviv University

REFERENCES

American Psychological Association. 2010. *Publication Manual* (Appendix - pp. 245–254). American Psychological Association: Washington, DC.

Apicella, C. L., Carre, J. M., & Dreber, A. 2015. Testosterone and economic risk taking: A review. *Adaptive Human Behavior and Physiology*, 1: 358–385.

Bamberger, P., & Ang, S. 2015. The quantitative discovery: What it is how and how to get it published. *Academy of Management Discoveries*, 2: 1–6.

Bettis, R. A. 2012. The search for asterisks: Compromised statistical tests and flawed theory. *Strategic Management Journal*, 33: 108–113.

Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. 2016a. Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*, 37: 257–261.

Bettis, R. A., Helfat, C. E., & Shaver, J. M. 2016b. The necessity, logic, and forms of replication. *Strategic Management Journal*, 37: 2193–2203.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. T. 2009. *Introduction to meta-Analysis*. Chichester, UK: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. T. 2013. *Meta-regression manual*. Englewood, NJ: Biostat, Inc.

Brandt, M. J., et al. 2014. The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50: 217–224.

Carney, D. R., & Mason, M. F. 2010. Moral decision making and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology*, 46: 668–671.

D'Innocenzo, L., Luciano, M. M., Mathieu, J. E., Maynard, M. T., & Chen, G. 2016. Empowered to perform: A multilevel investigation of the influence of empowerment on performance in hospital units. *Academy of Management Journal*, 59: 1290–1307.

Eden, D. 2002. Replication, meta-analysis, scientific progress, and *AMJ's* publication policy. *Academy of Management Journal*, 45: 841–846.

Frone, M. 2013. *Alcohol and illicit drug use in the workforce and workplace*. Washington, DC: American Psychological Association.

Ganzach, Y. 2016. Cognitive ability and party identity: No important differences between democrats and republicans. *Intelligence*, 58: 18–21.

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. 2016. Comment on "Estimating the reproducibility of psychological science". *Science*, 351: 1037.

Hedges, L. V., & Olkin, I. 1985. *Statistical methods for meta-analysis*. New York: Academic Press.

Infurna, F. J., & Luthar, S. S. 2016. Resilience to major life stressors is not as common as thought. *Perspectives on Psychological Science*, 11: 175–194.

Ioannidis, J. P. A. 2005. Why most published research findings are false. *PLoS Medicine*, 2: 696–701.

Junni, P., Sarala, R. M., Taras, V., & Tarba, S. Y. 2013. Organizational ambitexterity and performance: A meta-analysis. ***Academy of Management Perspectives***, 4: 299–312.

Lipsey, M. W., & Wilson, D. B. 1993. The efficacy of psychological, educational, and behavioral treatment. ***American Psychologist***, 48: 1181–1209.

Lipsey, M. W., & Wilson, D. B. 2001. ***Practical meta-analysis***. Thousand Oaks, CA: Sage.

McKee, R. A., & Miller, C. C. 2015. Institutionalizing Bayesianism within the organizational sciences: A practical guide featuring comments from eminent scholars. ***Journal of Management***, 41: 471–490.

Meschi, P.-X., Metais, E., & Miller, C. C. 2015. Leader longevity, cognitive inertia, and performance in organizations with stretch goals: Evidence from "La Royale" and its ambition to gain naval supremacy between 1689 and 1783. In G. Gavetti & W. Ocasio (Eds.) ***Advances in strategic management: Cognition and strategy,*** vol. 32: 467–504. Bingley, UK: Emerald Group Publishing Limited.

Miller, C. C., & Cardinal, L. B. 1994. Strategic planning and firm performance: A synthesis of more than two decades of research. ***Academy of Management Journal***, 37: 1649–1665.

Mintzberg, H. 1994. ***The rise and fall of strategic planning***. New York: Free Press.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science, ***Science***, 349: 1–8. doi:dx.doi.org/10.1126/science.aac4716.

Riskin, A., et al. 2015. The impact of rudeness on medical team performance: A randomized trial. ***Pediatrics***, 136: 487–495.

Rosenthal, R., & Rubin, D. B. 1982. A simple, general purpose display of magnitude of experimental effect. ***Journal of Educational Psychology***, 74: 166–169.

Schmidt, F. L., & Hunter, J. E. 2015. ***Methods of meta-analysis: Correcting error and bias in research findings*** (3rd ed.). Thousand Oaks, CA: Sage.

Schmidt, F. L., & Oh, I.-S. 2016. The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? ***Archives of Scientific Psychology***, 4: 32–37.

Schneider, B., & Bowen, D. E. 1985. Employee and customer perceptions of service in banks: Replication and extension. ***Journal of Applied Psychology***, 70: 423–433.

Starbuck, W. H. 2006. ***The production of knowledge: The challenges of social science research***. Oxford, U.K.: Oxford University Press.

Van de Ven, A. H. 2016. Happy Birthday, *AMD*. ***Academy of Management Discoveries***, 2: 1–3.