

2018

Place of behavior analysis in the changing culture of replication and statistical reporting in psychological science

Abdulrazaq Imam

John Carroll University, aimam@jcu.edu

Follow this and additional works at: https://collected.jcu.edu/fac_bib_2018


 Part of the [Applied Behavior Analysis Commons](#)

Recommended Citation

Imam, Abdulrazaq, "Place of behavior analysis in the changing culture of replication and statistical reporting in psychological science" (2018). *2018 Faculty Bibliography*. 34.
https://collected.jcu.edu/fac_bib_2018/34

This Article is brought to you for free and open access by the Faculty Bibliographies Community Homepage at Carroll Collected. It has been accepted for inclusion in 2018 Faculty Bibliography by an authorized administrator of Carroll Collected. For more information, please contact connell@jcu.edu.

Place of behavior analysis in the changing culture of replication and statistical reporting in psychological science

Abdulrazaq A. Imam 

Department of Psychological Science, University Heights, OH, USA

ABSTRACT

The “new” psychological science seeks to promote a culture of replication in response to rampant publication bias and some controversial failures to replicate. Two of the solutions adopted are the emphases on the New Statistics by *Psychological Science* and the growing use of replication repositories. Where does behavior analytic research fit in this effort? Although experimental and applied research methods in behavior analysis naturally are replication focused and replication friendly, trending growth in group designs in different areas of behavior analytic research suggests that we should be sharing the concerns in the broader psychological science community. What measures, if any, are behavior analytic journals taking or should be taking to address these concerns? Not all of the solutions under consideration in the “new” psychological science are amenable to behavior analytic research. How do we proceed? Recommendations include formulating editorial policies in behavior analytic journals and organizational programming collaborations, for example, between Association for Behavior Analysis International and the Association for Psychological Science.

KEYWORDS

Professional organizational collaboration; psychological science; behavior analysis; editorial policies; publication bias; estimation; preregistration

Since its introduction from research in agriculture and wide acceptance in psychology a little before the middle of the last century (Rucci & Tweney, 1980), the use of null hypothesis statistical testing (NHST) in psychological research has caught on like a wild fire without bounds. Over the years, many have pointed out how problematic and detrimental the practice is for the science of psychology, but with little heed. Recent developments in the world of science at large (e.g., Fidler, Thomason, Cumming, Finch, & Leeman, 2004), and in psychological science in particular, have rekindled interest and brought the issues to the fore. The present commentary explores the nature of the problem within psychological science, the current efforts to proffer solutions to those problems, and the role behavior analysis may play in fostering or benefitting from those efforts, broadly and specifically.

The problem

Various efforts have been underway in the larger world of psychological science to address the tripartite problem of a poor replication culture, of p -hacking resulting from the heavy reliance on NHST, and of the pervasive publication bias associated with it (Pashler & Wagenmakers, 2012). The replication problem has two dimensions: (1) the almost complete absence of the practice of replicating important findings in the field, which may be a byproduct of the dominance of NHST in the ubiquitous use of group designs in psychological research, and (2) the high-profile failures that have been recorded in such areas as priming (Cesario, 2014; Dijksterhuis, 2014; Klatzky & Creswell, 2014; Simons, 2014; Stroebe & Strack, 2014) and precognition (Alcock, 2011; Bem, 2011; Ritchie, Wiseman, & French, 2012).

The problem of p -hacking is peculiar to NHST in group designs and manifests when researchers attempt to unsystematically achieve $p < .05$ by adding participants to groups after the fact, solely because, statistically, the larger the N , the more likely that p will approach 0. The reason most do it, of course, is that publication outlets would not accept reports of $p > .05$ that is considered statistically not significant, which inherently leads to publication bias because only statistically significant findings see the light of day. By publishing only statistically significant research outcomes, the literature is awash with only positive findings that result in a lopsided body of knowledge, which is antithetical to building an integrative, cumulative science. The presenting challenge, of course, is the combination of the tripartite problems of irreproducibility of findings, p -hacking, and publication bias (Ioannidis, 2005; Simmons, Nelson, & Simonsohn, 2011). Of course, these problems are not peculiar to psychology by any means. As prevalent as they are in other sciences, however, they affect psychological science in ways that psychologists can no longer afford to ignore them, to the extent that we seek to build a cumulative science (Branch, 2014; Meehl, 1978).

Current efforts

One could argue that organizational support for efforts to address the tripartite problems among bodies representing vast numbers of psychologists has been varied in extent and intensity. For example, the American Psychological Association (APA) has supported publications such as Kline (2013) that is now in its second edition and has advocated for the use of effect size in statistical reporting in the current edition of the APA Publication Manual (American Psychological Association, 2009). As far as advocacy goes, however, that is about it. In contrast, the Association for Psychological Science (APS) has devoted major volumes in its publication outlets including the *Psychological Science (PS)* and *Perspectives on Psychological Science (PoPS)* to major discussions of the issues involved and has begun, in earnest, to advocate more rigorously and actively for changes in the practice of the science. What is different also in the latter efforts is that the APS has gone beyond mere advocacy to implementation, the key ingredient lacking in previous attempts, and in admonitions of the use and abuse of NHST with the attendant problems of replications and replicability, on the one hand, and publication bias, on the other. These problems are intractable for the thriving science that psychology aspires to become.

What role for behavior analysts?

I recall in my graduate school days the debates that accompanied concerns for the future of behavior analysis in the larger discipline of psychology with solutions proffered ranging from praxics to behaviorology, both representing an actual name change (Epstein, 1984; Fraley, 1987; see Malagodi & Branch, 1985) in contrast to paradigmatic behaviorism (Staats, 1986). For the most part, behavior analysis has remained mostly in psychology and education departments across the United States and in many parts of the world. Many of the points of contention in the discussion centered on how behavior analysis and psychology differ from one another, in terms of philosophy, methodology, and general goals and objectives. Besides keeping the name “behavior analysis” within psychology (a la Division 25 of the APA), recent developments in the area proffered substantial progress, the least of which is not a thriving certification of practitioners in applied behavior analysis started about a decade ago (BACB). Whereas the calls for separation may have failed, in practice, behavior analysis and psychology have remained relatively isolated from each other, a marriage of convenience in which each is blissfully doing its own thing, for the most part, with limited “cross-pollination”. The more recent wave of interest in the methodological concerns of replicability, *p*-hacking, and publication bias led by the APS represents a development within psychology in which behavior analysis may be in a position to simultaneously contribute as well as benefit from the proffered solutions. Broadly, the solutions being pursued currently to address these problems include promoting the adoption of the “new statistics” (Cumming, 2014a) and encouragement of various replication efforts (Pashler & Wagenmakers, 2012) in the field. Why should behavior analysts care about these efforts?

To start with, in recent decades, there has been growing reports of NHST in the pages of journals that publish behavior analytic research, including *Journal of the Experimental Analysis of Behavior (JEAB)* and *The Psychological Record (TPR)*. Zimmermann, Watkins, and Poling reported “progressive increase over time in the percentage of articles reporting an inferential statistic” (2015, p. 209) in *JEAB* involving both human and nonhuman subjects. Such development is partly due to the growth in topical coverage within behavior analysis; for example, it is not uncommon to see reports on stimulus equivalence relying on NHST. The “new statistics” recommends the estimation approach, which requires reporting of effect sizes (ESs), confidence intervals (CIs), and meta-analyses, over NHST. The recommendation is in agreement with those of the current APA publication manual (American Psychological Association, 2009), and psychological journals, particularly those published by the APS, have begun requiring it (e.g., Eich, 2014; *Psychological Science*, 2017). What would *European Journal of Behavior Analysis (EJOBA)*, *Journal of Applied Behavior Analysis (JABA)*, *JEAB*, *Mexican Journal of Behavior Analysis (MJBA)*, *The Behavior Analyst (TBA)*, and *TPR* do? Are these behavior analytic journals going to begin to require reports of ESs, CIs, and meta-analyses in studies reporting group-design data (see Parker, Vannest, & Davis, 2011 for Small-*N* alternatives) as other psychological journals such as *PS* have started to do in accordance with the recommendations of the latest version of the APA manual (American Psychological Association, 2009)? There are good reasons to suggest that we should. A cursory check of author guidelines and instructions to authors in the latest volumes of these behavioral journals, including *Behavioural Processes (BP)*, however, did not include standards for statistical reporting like the ones provided by *PS* (e.g., reporting CIs instead

of standard deviation [SD] or SE on point estimates), despite their growing use (e.g., Zimmermann et al., 2015). *PS* provides authors with a recommendation and a rationale for it, and specifies formatting standards (e.g., reporting exact *p*-values instead of relative *p*-values as is commonly done), in addition (see *Psychological Science*, 2017, for further details).

Another corollary of these efforts that should interest behavior analysts is what Cumming (2014b; 7:00–11:30; 2014c; 26:50) refers to as “statistical cognition”. According to Cumming, because there has been little research on the use of NHST by researchers in various scientific fields in the literature, psychologists are best positioned to conduct research on how they use and interpret data. His own team’s research (e.g., Coulson, Healey, Fidler, & Cumming, 2010) examining how people “think, talk, and feel about *p*-value” (Cumming, 2014b, 11:42) illustrates such effort. One of the major findings is that despite seeing only point estimates with their CIs, some researchers still referenced the data in terms of NHST! Cumming noted that even when they saw CIs they *think* of NHST (my emphasis), concluding that perhaps we might be better off doing away with reporting *p*-values altogether when reporting CIs. In fact, however, the data supporting that conclusion were only *mentions* of NHST, not what the researchers *thought* (see Cumming, 2014b, 8:44–10:00; my emphases) about NHST. Again, there is a plenty of opportunity for behavior analysts to contribute to such efforts to improve our understanding of scientific *behavior*.

Indeed, Cumming himself appeared to have provided such framework in his presentation, perhaps without meaning to do so. He characterized people’s relationship to *p*-values in terms of how they “think, talk, and feel” about it (Cumming, 2014b, 12:38–14:39). I encourage readers to watch the segment. In a nutshell, it reads something like this: (1) *p* is a measure of strength (a la Fisher); (2) ... that *elicits* significance language; (3) ... which *suggests* truth; (4) ...*evokes* emotions; and (5) has real-life *consequences!* (my emphases, but the exclamation is in the original). These are familiar references to this readership. What contingencies support the widespread use of NHST, not only in the wider psychological science community, but also increasingly in behavior analysis, despite its shortcomings? What remedies are necessary to tackle adequately and effectively the attendant problems? These and many other related questions are well within the purview of behavior analysis; the potential is vast and the ground fertile. Smaldino and McElreath (2016) offered a recent natural selection perspective on doing bad science. Surely, others closer to home based on selection by consequences (Skinner, 1981) can be rendered. Incidentally, one of the strategies adopted in *PS* is to assign three different Open Science Forum (OSF) badges for published articles meeting criteria for preregistration, open materials, and open data, in their promotion of replication efforts (Cumming, 2014c; 19:55; see also *Journal of Experimental Social Psychology*, 2018; for their adoption). Such badges appear to be veritable consequences! What are the implications of providing such powerful consequences for something like preregistration that is antithetical to behavior analytic research (see further discussion below; see also Koole & Lakens, 2012)? Could alternatives be developed that would meet the needs of behavior analytic approaches? These are additional reasons for behavior analysts to get involved and embrace these efforts.

Furthermore, of the many efforts already underway on statistical reform promoting the estimation approach outlined by Cumming (2014c), one is particularly noteworthy in support of behavior analysts’ involvement. Geoff Loftus was the editor for *Memory*

and Cognition from 1993 to 1997 during which he promoted use of error bars on graphs. Cumming reported a decline in the practice after his editorial term ended (2014c, 21:26; see Finch et al., 2004). The lesson here is that contingency management, as we know, requires effective managing for long-term maintenance of behavior; the rules are not enough. We know a thing or two about rule-governed behavior (Hayes, 1989). As behavior analysts, how can we advance sound reporting practices in psychological science? Various projects on research integrity also are underway on the related matter of transparency in support of the replication efforts. A couple of them may have direct implications for behavior analysis, namely, the OSF under the Center for Open Science and the Registered Replication Report of *PoPS* (Cumming, 2014c; 14:06–15:48; see *Psychological Science*, n.d.). The problem is that they require a declaration of protocols and/or preregistration, both of which may be antithetical to the way we tend to conduct experiments in behavior analysis (Neuringer, 1991; Sidman, 1960). The issue is that given the rigor and painstaking data collection that is characteristic of Small-*N* designs commonly used in behavior analysis that render it “slow” compared to the “snapshot” approach of group designs that render them “fast”, the first provides a leaner “schedule” than the second one does as far as building careers in psychological science is concerned. Consequently, one could argue that psychologists have been *differentially reinforced* for using group designs, and for all the attendant *p*-hacking, more than for using Small-*N* designs! This is a challenge for behavior analysts in the new efforts, requiring further analysis and attention. Part of that analysis would have to consider how selective attention to novel findings (publication bias) in psychological science has promoted selection for novelty, which obviates focus on replication. Replication, of course, is inherently integral to the various Small-*N* designs commonly used in behavior analysis (Branch, 2014; Sidman, 1960). Interestingly, even *PS* cannot seem to wean itself, despite APS’s advocacy efforts to the contrary, of the root of this publication bias: “Preference is given to papers that make a new and notable contribution—an idea, a discovery, a connection—to psychological science, broadly interpreted to include emerging as well as established areas of research” (*Psychological Science*, 2017).

Cumming (2014d, 27:40) concluded his remarks on the ongoing APS efforts with an eight-step outline of how to implement the estimation approach in the new psychology. These warrant comments to provide a framework for how behavior analysts may proceed in getting involved in the ongoing efforts in the wider psychological science community. The first step seeks to redirect psychologists from testing null by asking “is this treatment better” (Cumming, 2014a, p. 14) to adopting estimation by asking “How much ...?”, “To what extent ...?”, “How many ...?”, which he considers “the key to a more quantitative discipline” (Cumming, 2014b, 27:40). The estimation alternative offered in place of NHST excludes the “What if ...?” or “I wonder what would happen if ...” sort of question that might precede a hunch as part of “scientific play” (Neuringer, 1991, pp. 7–8) common in behavior analytic research (Sidman, 1960). What is the implication of this gap for the place of behavior analysis in this new culture in psychological science? The second step simply advocates matching the choice of ESs to the research question. The third step, however, seeks a complete statement of intent concerning “procedure, data analysis”, etc. (Cumming, 2014b; 27:40; see also *Psychological Science*, n.d.). This could be highly problematic from a behavior-analytic perspective such as that provided by Neuringer. Additionally, according to Sidman, “there

can be no explicit rules for determining the most appropriate replicative technique. The method to be employed will be selected from the choices made available by the experimental data and by the control techniques at hand” (1960, p. 138). Prespecification seems to preclude adopting techniques guided by the experimental data.

The fourth step urges psychologists to compute and report “point and interval estimates (CIs) for [the chosen] ESs”, and the fifth advocates graphing error bars with CIs (Cumming, 2014b, 27:40). Both of these are useful for behavior analysis as well and warrant no further commentary. Step 6 “calls for informed judgement, rather than a mechanistic statement of statistical significance” (Cumming, 2014a, p. 14). This is a welcome but familiar process for most behavior-analytic researchers. We are in a better position to contribute substantially and substantively, given the differential contingency placed on use of group designs mentioned above, with our experiences in relying on “subjective” judgments common in visual inspection of data, for example. As Sidman put it

Whether or not we make an inductive inference, and the degree of tenacity with which we cling to that inference, will depend upon our behavioral history (experience). I refer to this history when I say that the evaluation of generality is a matter of *judgment*. From an act of induction based upon our own accumulated experience, we judge the amount of generality to be added to a variable when it proves effective in experiments that have little or no operational connection with each other. (1960, p. 59; my emphasis)

The kind of judgments involved in estimation is one that is unfamiliar to most psychologists because of the heavy reliance on dichotomous thinking inherent in NHST. Alternative ways of thinking about data and analyses (e.g., Branch, 2014; Fiedler, 2017; Killeen, 2005a, 2005b; Sidman, 1960) may be ripe for consideration or reconsideration by the wider psychological science community. Behavior analysts are positioned to play a beneficial role. The final two steps call for meta-analytic thinking and complete reporting, both of which are useful for behavior analysis as well. Altogether, these steps seem amenable segways to having an important impact on the process that is underway in the current culture changes in psychological science.

Finally, the foregoing is not meant to suggest that there has been no involvement by behavior analysts at all in the debates concerning the problematic excessive reliance on NHST by psychologists. Notable exceptions are Branch (1999, 2014) and Killeen (2005a, 2005b, 2006), both of whom have made substantive contributions to the debates (Cumming, 2005; Doros & Geier, 2005; Macdonald, 2006; Wagenmakers & Grunwald, 2006) over the status and potential solutions to the problems associated with the pervasive use of NHST in psychological science. What has yet to happen thus are organizational and editorial policy movements in the Association for Behavior Analysis International (ABAI) and allied parties like the Experimental Analysis of Behavior Group and in behavior-analytic journals regarding the ongoing efforts in the extant psychological science community. ABAI and APS have opportunities to collaborate and coordinate. Every now and then, for example, they hold their annual meetings in Chicago during the Memorial Day weekend but never coordinate programming. These are missed opportunities. Furthermore, the APS also has begun an international convention of psychological science that has been so far held in European countries; the next is in Paris, France, in March 2019. Could EAHB and APS hold meetings around the same times and locale to allow for

fruitful collaborations and exchanges, given that the two organizations hold their meetings in spring? Joint or shared workshops could serve as springboards for educating members of the respective organizations on methodological concerns that could shed light on common threads and debate points of methodological contention across perspectives. What measures, if any, are behavior analytic journals taking or should be taking to address the changes in statistical reporting practices (e.g., Cumming et al., 2007) that are being suggested in the ongoing efforts? A good place to start is adopting editorial policies on statistical reporting practices similar to those of *PS* (see *Psychological Science*, 2017) so that they can begin to move away from discredited NHST practices and behavior analytic authors using group designs can begin to update their analytic and reporting standards.

Concluding remarks

In a recent study comparing reporting practices of two behavioral journals in psychological science, *JEAB* and *BP*, in 2011 and 2015 (the years before and after the publication of the *PoPS* papers on the change advocacy); for example, the authors used error bars on graphs in both journals, but *BP*'s use was higher than *JEAB*'s. Both journals reported standard deviation (*SD*) and standard error of the mean (*SEM*) in both years, but *BP* reported more *SEM* than *SD* in both years; only in 2015 was this true for *JEAB*. Reporting of *CI*, unfortunately, was very limited in *JEAB* in both years compared to *BP*. The only *ES* measures reported by *JEAB* were partial η^2 and Cohen's *d* in both years; both measures, as well as reporting of η^2 , were higher in 2015 than 2011 in *BP*. Both journals also reported actual replications in both years, although *JEAB* recorded a decline in 2015 despite higher reports than *BP* in both years (Frate & Imam, 2016). Thus, whereas these behavioral journals unsurprisingly reported replications (see above; also Branch, 2014), they were not consistent universally across the board on statistical reporting practices as would be expected under the estimation approach of the new psychological science. As noted above, therefore, it will be useful to have coherent editorial policies and guidelines for authors on statistical reporting standards across behavior analytic journals. Perhaps these and other efforts would provide a conduit to greater recognition of the *presence* and *place* of behavior analysis in the new and growing culture of replication and statistical reporting practices in psychological science, to contrast well with the debates of the 1980s and 1990s about where we stood with respect to the rest of psychology.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Abdulrazaq A. Imam  <http://orcid.org/0000-0002-1262-6022>

References

- Alcock, J. (2011, January 6). Back from the future: Parapsychology and the Bem affair. Center for Skeptical Inquiry. Retrieved from http://www.csicop.org/specialarticles/show/back_from_the_future
- American Psychological Association. (2009). *Publication Manual of the American Psychological Association* (6th ed.). Washington, D. C.: Author.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*, 407–425.
- Branch, M. N. (1999). Statistical inference in behavior analysis: Some things significance testing does and does not do. *The Behavior Analyst, 22*, 87–92.
- Branch, M. N. (2014). Malignant side effects of null-hypothesis significance testing. *Theory & Psychology, 24*, 256–277.
- Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science, 9*, 40–48.
- Coulson, M., Healey, M., Fidler, F., & Cumming, G. (2010). Confidence intervals permit, but do not guarantee, better inference than statistical significance testing. *Frontiers in Quantitative Psychology and Measurement, 1*, 1–9.
- Cumming, G. (2005). Understanding the average probability of replication. *Psychological Science, 16*, 1002–1004.
- Cumming, G. (2014a). The new statistic: Why and how? *Psychological Science, 25*, 7–29.
- Cumming, G. (2014b, May 22). The new statistics: Estimation and research integrity (PART 1: Confidence intervals, NHST, and p values) [video file]. Retrieved from <https://www.psychologicalscience.org/members/new-statistics>
- Cumming, G. (2014c, May 22). The new statistics: Estimation and research integrity (Part 2: Research integrity and the new statistics) [video file]. Retrieved from <https://www.psychologicalscience.org/members/new-statistics>
- Cumming, G. (2014d, May 22). The new statistics: Estimation and research integrity (Part 6: Meta-Analysis and Meta-Analytic Thinking) [video file]. Retrieved from <https://www.psychologicalscience.org/members/new-statistics>
- Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., ... Wilson, S. (2007). Statistical reform in psychology: Is anything changing? *Psychological Science, 18*, 230–232.
- Dijksterhuis, A. (2014). Welcome back theory! *Perspectives on Psychological Science, 9*(72–75). doi:10.1177/1745691613513472
- Doros, G., & Geier, A. B. (2005). Probability of replication revisited. *Psychological Science, 16*, 1005–1006.
- Eich, E. (2014). Business not as usual. *Psychological Science, 25*, 3–6.
- Epstein, R. (1984). The case for praxics. *The Behavior Analyst, 7*, 101–119.
- Fidler, F., Thomason, N., Cumming, G., Finch, S., & Leeman, J. (2004). Editors can lead researchers to confidence intervals, but can't make them think: Statistical reform lessons from medicine. *Psychological Science, 15*, 119–126.
- Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a prior theorizing. *Perspectives on Psychological Science, 12*, 46–61.
- Finch, S., Cumming, G., Williams, J., Palmer, L., Griffith, E., Alders, C., ... Goodman, O. (2004). Reform of statistical inference in psychology: The case of *Memory & Cognition*. *Behavior Research Methods, Instruments, & Computers, 36*, 312–324.
- Fraley, L. E. (1987). The cultural mission of behaviorology. *The Behavior Analyst, 10*, 123–126.
- Frate, M., & Imam, A. A. (2016, May). *Replication and statistical reporting practices in two behavioral journals*. Poster presented at the 28th annual meeting of the Association for Psychological Science, Chicago, IL.
- Hayes, S. C. (Ed.). (1989). *Rule-governed behavior: Cognition, contingencies, and instructional control*. Reno, NV: Context Press.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*, 0696–0701.

- Journal of Experimental Social Psychology. (2018, February 18). Guide for authors. Retrieved from <https://www.elsevier.com/journals/journal-of-experimental-social-psychology/0022-1031/guide-for-authors>
- Killeen, P. R. (2005a). An alternative to null-hypothesis significance tests. *Psychological Science*, *16*, 345–353.
- Killeen, P. R. (2005b). Replicability, confidence, and priors. *Psychological Science*, *16*, 1009–1012.
- Killeen, P. R. (2006). The problem with Bayes. *Psychological Science*, *17*, 643–644.
- Klaczky, R. L., & Creswell, J. D. (2014). An intersensory interaction account of priming effects—And their absence. *Perspectives on Psychological Science*, *9*, 49–58.
- Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). Washington, D. C.: American Psychological Association.
- Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Psychological Science*, *7*, 608–614.
- Macdonald, R. R. (2006). Why replication probabilities depend on prior probability distributions. *Psychological Science*, *16*, 1007–1008.
- Malagodi, E. F., & Branch, M. N. (1985). Praxics and behaviorism. *The Behavior Analyst*, *8*, 123–125.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.
- Neuringer, A. (1991). Humble behaviorism. *The Behavior Analyst*, *14*, 1–13.
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*, 303–322.
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*, 528–530.
- Psychological Science. (2017, October 30). Submission guidelines. Retrieved from https://www.psychologicalscience.org/publications/psychological_science/ps-submissions
- Psychological Science. (n.d.). Preregistration of research plans. Retrieved from https://www.psychologicalscience.org/publications/psychological_science/preregistration
- Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: Three unsuccessful attempts to replicate Bem's retroactive facilitation of recall effect. *PLoS ONE*, *7*(3), e33423.
- Rucci, A. J., & Tweney, R. D. (1980). Analysis of variance and the "second discipline" of scientific psychology: A historical account. *Psychological Bulletin*, *87*, 166–184.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Boston, MA: Authors Cooperative.
- 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.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76–80.
- Skinner, B. F. (1981). Selection by consequences. *Science*, *213*, 501–504.
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*, 1–17.
- Staats, A. W. (1986). Left and right paths for behaviorism's development. *The Behavior Analyst*, *9*, 231–237.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*, 59–71.
- Wagenmakers, E.-J., & Grunwald, P. (2006). A Bayesian perspective on hypothesis testing. *Psychological Science*, *16*, 641–642.
- Zimmermann, Z. J., Watkins, E. E., & Poling, A. (2015). *JEAB* research over time: Species used, experimental designs, statistical analyses, and sex of subjects. *The Behavior Analyst*, *38*, 203–218.