PSYCHOPHYSIOLOGY

Psychophysiology, 54 (2017), 3–5. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/psyp.12787

Introduction to the special issue on recentering science: Replication, robustness, and reproducibility in psychophysiology

EMILY S. KAPPENMAN[a] AND ANDREAS KEIL[b]

[a]Department of Psychology, San Diego State University, San Diego, California, USA
[b]Department of Psychology and Center for the Study of Emotion and Attention, University of Florida, Gainesville, Florida, USA

Abstract

In recent years, the psychological and behavioral sciences have increased efforts to strengthen methodological practices and publication standards, with the ultimate goal of enhancing the value and reproducibility of published reports. These issues are especially important in the multidisciplinary field of psychophysiology, which yields rich and complex data sets with a large number of observations. In addition, the technological tools and analysis methods available in the field of psychophysiology are continually evolving, widening the array of techniques and approaches available to researchers. This special issue presents articles detailing rigorous and systematic evaluations of tasks, measures, materials, analysis approaches, and statistical practices in a variety of subdisciplines of psychophysiology. These articles highlight challenges in conducting and interpreting psychophysiological research and provide data-driven, evidence-based recommendations for overcoming those challenges to produce robust, reproducible results in the field of psychophysiology.

Descriptors: Reproducibility, Replicability, Psychophysiology, Psychometrics

From its origins in the 1940s, the multidisciplinary field of psychophysiology has engaged in a systematic analysis of mind-body interactions using a range of ever-advancing research techniques. At present, psychophysiologists employ autonomic nervous system measures; electrical, magnetic, optical, and hemodynamic measures of brain activity; indices of endocrine processes; electromyography; and many other measures of bodily processes. In the first issue of the journal *Psychophysiology*, then Editor-in-Chief Albert Ax emphasized the need for a multidisciplinary approach in connecting physiology with psychological processes (Ax, 1964): "Modern psychophysiology is a response to the challenge inherent in the full realization of the complex nature of the human organism. The concepts of homeostasis, cybernetics, information and systems theory together with the facts of physiology and psychology define the challenge." To date, psychophysiology has maintained and further developed this focus on sophisticated, multimethod, and multidiscipline approaches. Accordingly, the field has been defined to a large extent by efforts to develop, compare, and evaluate methods—an endeavor that has long been encouraged and supported by *Psychophysiology*, the flagship journal in the field (see, e.g., Fabiani, 2015, for a detailed discussion of this topic).

The authors would like to thank the researchers who contributed to this special issue. The support of Monica Fabiani, editor-in-chief of *Psychophysiology*, is highly appreciated. Work on this issue was partly supported by a grant from the National Institutes of Mental Health (R01MH097320) to AK.

Address correspondence to: Emily S. Kappenman, Department of Psychology, San Diego State University, 6363 Alvarado Court, Suite 250, San Diego, CA 92120, USA. E-mail: Emily.Kappenman@sdsu.edu

The goal of the present special issue is to build on this long-held custom in psychophysiology of applying rigor and systematic evaluation to our own field. This emphasis is particularly timely given the current discussion of a "replicability crisis" (Pashler & Harris, 2012), which has included debate about surprisingly high numbers of positive results (Ioannidis, 2005), underpowered studies (Button et al., 2013), and most recently by large-scale studies reporting failures to replicate published findings in fields such as psychology (Open Science Collaboration, 2015), medicine (Begley & Ellis, 2012), and neuroscience (Steckler, 2015).

The response to the replicability crisis has varied by discipline but has generally included a surge in articles providing guidelines and recommendations for best practices, with the intention of improving scientific rigor and enhancing reproducibility of empirical findings within and between laboratories. In a similar vein, new guidelines from the U.S. National Institutes of Health emphasize rigor when evaluating grant submissions. Although the overall goals of these initiatives are generally agreed upon, there has been intense debate regarding the nature and extent of such mandated changes to current scientific practice. The urgency and productivity of this debate is well illustrated by the fact that the space allotted for this editorial does not allow us to present even a condensed list of all the procedures proposed by different authors, scientific organizations, journals, interest groups, and funding agencies.

For example, several journals have updated their policies to require explicit justification of sample size selection through power analyses (Lindsay, 2015), and to discourage traditional null-hypothesis testing in favor of what is referred to as the "new statistics" (Eich, 2014). Similarly, many authors of guidelines and best practices articles provide encouragement to use larger samples and avoid underpowered studies (Button et al., 2013). Many outlets now require or strongly encourage that all data and analysis scripts are made openly accessible (Barbui, 2016), and some have recommended that hypotheses and methods be preregistered before data collection commences (Dal-Ré et al., 2014). These measures are intended to discourage a host of behaviors, such as so-called *p*-hacking (Simonsohn, Nelson, & Simmons, 2014) and post hoc adjustment of hypotheses (Francis, 2013). Recommendations have also been made regarding the presentation of data in figures, aiming to improve the ability to visualize variability and consistency rather than merely central tendency (Allen, Erhardt, & Calhoun, 2012). Not surprisingly, changes to the way we conduct research training in ethics (Hauser, 2014) and statistical methods (Cumming, 2014) have been called for as well. Many of these initiatives aim to change policies and establish new standards, with the goal of improving scientific practices in the community as a whole, and to facilitate detection and correction of reports presenting data not considered robust or replicable.

Although establishing general guidelines can provide certain benefits, the journal *Psychophysiology* has long embraced a different approach, which focuses on rigorously evaluating methods and procedures in an effort to provide field-specific, data-driven guidelines, in contrast to broad policies. The present special issue builds on this existing tradition of psychophysiology as a rigorous, systematic science.

Instead of proposing broad policies or establishing general methodological or statistical standards, the articles in this special issue provide concrete steps that psychophysiologists can take toward a more rigorous, systematic program of research. Specifically, data-driven recommendations are provided for designing and planning studies, analyzing complex psychophysiological data sets, interpreting results, and communicating findings to the scientific community that will heighten the reproducibility of published reports in our field.

Overview of Special Issue Articles

Providing a conceptual and historical framework for the special issue, the lead article by Margaret Bradley discusses the evolution of funding and publishing in science over the course of her career, specifically with regard to the current focus on obtaining "novel" and "breakthrough" results and the systematic de-emphasis of incremental, programmatic research **(Bradley, 2017).**

Several contributions systematically evaluate the psychometric properties of existing tasks, measures, materials, and/or analysis approaches in a variety of subdisciplines of psychophysiology, with sometimes surprising results:

Hess and colleagues (2017) examine the internal reliability and long-term stability of surface facial electromyography in two widely used tasks and demonstrate the complexity of people's facial reactions to affective stimuli.

Boekel, Forstmann, and Keuken (2017) examine the test-retest reliability of measures of white matter tracts derived from diffusion tensor imaging and show that the stability of the measures decreases as the temporal distance between measurements increases.

Tenke and colleagues (2017) measured test-retest reliability of electrophysiological markers for predicting response to antidepressants and show good-to-excellent reliability across different laboratories.

Kuntzelman and Miskovic (2017) examine the reproducibility of burgeoning graph theoretical measures as a way of characterizing network structure.

Maheux and colleagues (2017) show how the jackknife approach applied to functional near-infrared spectroscopy (fNIRS) can improve statistical power without inflating Type I error rate.

Mathewson, Harrison, and Kizuk (2017) provide a systematic comparison of new technologies available in EEG electrodes, including active electrodes with online amplification and dry (gel-free) electrodes, with traditional materials.

Drisdelle, Aubin, and Jolicoeur (2017) use simulated data to investigate the use of independent component analysis for ocular correction of horizontal eye movements in the context of experiments using lateralized ERP components.

Brooks, Zoumpoulaki, and Bowman (2017) show that the use of existing data in choosing regions of interest for analysis can in some circumstances be used without biasing results.

Meyer and colleagues (2017) provide an entirely new approach to examining the relationship between individual differences and psychophysiological measures through the use of a regression-based analysis as an alternative to traditional subtraction-based methods.

Finally, **Thigpen, Kappenman, and Keil (2017)** provide an example analysis for measuring and interpreting the internal consistency of commonly used measures in ERP research.

Two papers address pitfalls in statistical analysis and interpretation of complex, multidimensional psychophysiological data sets and present useful approaches to avoid these common statistical traps:

David Groppe illustrates the use of confidence intervals in providing bounds on the true size of an effect and provides freely available MATLAB software to make this approach available to researchers **(Groppe, 2017)**.

Luck and Gaspelin (2017) use real-world data to demonstrate the ease of obtaining spurious (i.e., not genuine) effects in psychophysiological experiments and describe strategies that can be used to avoid these problems.

Kaye, Bradford, and Curtin (2016) provide a comprehensive and detailed evaluation of the psychometric properties of both startle and corrugator responses in a large sample of participants, including systematic assessments of effect size, internal reliability, and test-retest reliability in three commonly used tasks in the field.

Concluding Remarks

Collectively, the papers in this special issue illustrate how a rigorous and systematic approach can provide insights into the richness of psychophysiological data sets and, ultimately, into the nature of mind-body interactions. We expect that many positive changes in current practices in psychophysiological research will follow from the systematic, methods-focused recommendations presented in this issue. Specifically, the present contributions demonstrate that the systematic evaluation of existing tasks, measures, equipment, and analysis approaches, as well as the development and testing of new methods, are central to the progression of psychophysiology as a science. We believe that such rigorous evaluation of methods should be a part of everyday research in psychophysiology, not just a topic for special issues in our field. We hope that the examples in this issue demonstrate the advantages of a programmatic, methods-centered approach, in contrast to imposing broad policies and inflexible guidelines for scientific practice as a response to the replicability crisis in science. We encourage psychophysiologists to consider how the quantitative evaluation of reliability and robustness may be incorporated into their own studies, as well as how their work may benefit from implementing the practical steps aimed at enhancing the reproducibility of scientific findings in psychophysiology that are provided in this special issue.

References

Allen, E. A., Erhardt, E. B., & Calhoun, V. D. (2012). Data visualization in the neurosciences: Overcoming the curse of dimensionality. *Neuron*, *74*(4), 603–608. doi: 10.1016/j.neuron.2012.05.001

Ax, A. F. (1964). Goals and methods of psychophysiology. *Psychophysiology*, *1*(1), 8–25. doi: 10.1111/j.1469-8986.1964.tb02616.x

Barbui, C. (2016). Sharing all types of clinical data and harmonizing journal standards. *BMC Medicine*, *14*, 63. doi: 10.1186/s12916-016-0612-8

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. doi: 10.1038/483531a

Boekel., W., Forstmann, B. U., & Keuken, M., C. (2017). A test-retest reliability analysis of diffusion measures of white matter tracts relevant for cognitive control. *Psychophysiology*, *54*(1), 24–33.

Bradley, M. (2017). The science pendulum: From programmatic to incremental and back? *Psychophysiology*, *54*(1), 6–11.

Brooks, J. L., Zoumpoulaki, A., & Bowman, H. (2017). Data-driven region-of-interest selection without inflating Type I error rate. *Psychophysiology*, *54*(1), 100–113.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. doi: 10.1038/nrn3475

Cumming, G. (2014). The new statistics why and how. *Psychological Science*, *25*(1), 7–29. doi: 10.1177/0956797613504966

Dal-Ré, R., Ioannidis, J. P., Bracken, M. B., Buffler, P. A., Chan, A.-W., Franco, E. L., . . . Weiderpass, E. (2014). Making prospective registration of observational research a reality. *Science Translational Medicine*, *6*(224), 1–4. doi: 10.1126/scitranslmed.3007513

Drisdelle, B. L., Aubin, S., & Jolicoeur, P. (2017). Dealing with ocular artifacts on lateralized ERPs in studies of visual-spatial attention and memory: ICA correction versus epoch rejection. *Psychophysiology*, *54*(1), 83–99.

Eich, E. (2014). Business not as usual. *Psychological Science*, *25*(1), 3–6. doi: 10.1177/0956797613512465

Fabiani, M. (2015). The embodied brain. *Psychophysiology*, *52*(1), 1–5.

Francis, G. (2013). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology*, *57*(5), 153–169. doi: 10.1016/j.jmp.2013.02.003

Groppe, D.M. (2017). Combating the scientific decline effect with confidence (intervals). *Psychophysiology*, *54*(1), 139–145.

Hauser, S. L. (2014). What ethics integration looks like in neuroscience research. *Annals of Neurology*, *75*(5), 623–624. doi: 10.1002/ana.24177

Hess, U., Arslan, R., Mauersberger, H., Blaison, C., Dufner, M., Denissen, J. J. A., & Ziegler, M. (2017). Reliability of surface facial electromyography. *Psychophysiology*, *54*(1), 12–23.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Med*, *2*(8), e124. doi: 10.1371/journal.pmed.0020124

Kaye, J. T., Bradford, D. E., & Curtin, J. T. (2016). Psychometric properties of startle and corrugator response in NPU, affective picture viewing, and resting state tasks. *Psychophysiology*, *53*(8), 1241–1255. doi: 10.1111/psyp.12663

Kuntzelman, K., & Miskovic, V. (2017). Reliability of graph metrics derived from resting-state human EEG. *Psychophysiology*, *54*(1), 51–61.

Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science*, *26*(12), 1827–1832. doi: 10.1177/0956797615616374

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157.

Maheux, M., Bisaillon-Sicotte, E., Tabrizi, S., Armony, J., L., Lina, J-M., & Jolicoeur, P. (2017). Optimal measurements of hemodynamic response latency in fNIRS using the jackknife approach. *Psychophysiology*, *54*(1), 62–73.

Mathewson, K. E., Harrison, T. J. L., & Kizuk, S. A. D. (2017). High and dry? Comparing active dry EEG electrodes to active and passive wet electrodes. *Psychophysiology*, *54*(1), 74–82.

Meyer, A., Lerner, M. D., De Los Reyes, A., Laird, R. D., & Hajcak, G. (2017). Considering ERP difference scores as individual difference measures: Issues with subtraction and alternative approaches. *Psychophysiology*, *54*(1), 114–122.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). doi: 10.1126/science.aac4716

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531–536. doi: 10.1177/1745691612463401

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. doi: 10.1037/a0033242

Steckler, T. (2015). Preclinical data reproducibility for R&D—The challenge for neuroscience. *Psychopharmacology*, *232*(2), 317–320. doi: 10.1007/s00213-014-3836-3

Tenke, C., A., Kayser, J., Pechtel, P., Webb, C. A., Dillon, D. G., Goer, F., . . . Bruder., G. E. (2017). Demonstrating test-retest reliability of electrophysiological measures for healthy adults in a multisite study of biomarkers of antidepressant treatment response. *Psychophysiology*, *54*(1), 34–50.

Thigpen, N. N., Kappenman, E. S., & Keil, A. (2017). Assessing the internal consistency of the event-related potential: An example analysis. *Psychophysiology*, *54*(1), 123–138.

(Received September 26, 2016; Accepted October 11, 2016)