International Journal of Psychophysiology 111 (2017) 5–16

Contents lists available at ScienceDirect

International Journal of Psychophysiology

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

Improving the rigor of psychophysiology research

Scott A. Baldwin *

Department of Psychology, Brigham Young University, United States

ARTICLE INFO

Article history:
Received 19 November 2015
Received in revised form 11 April 2016
Accepted 20 April 2016
Available online 29 April 2016

Keywords:
Replication
Rigor
Statistical power

ABSTRACT

Psychology as a field is in the midst of what is sometimes called a "crisis" because false findings are prevalent. Although the focus of the methodological and substantive criticisms of psychology has focused on social psychology, psychophysiology research is not without its problems. The author discusses (a) researcher flexibility and its impact on the stability of conclusions and (b) the role power plays in the probability that a finding is true and the precision of estimates. The author uses examples and data from psychophysiological research to illustrate the problems. The author concludes with a discussion of ways to shift the practice of science to improve the reliability of findings. Suggestions for improvement include: increased power through collaboration, improved statistical and methodological training, pre-registration of studies, improved reporting standards, and shifting incentives surrounding hiring and promotion.

> *There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims* (Ioannidis, 2005, p. 0696)

1. Introduction

Rigorous, replicable, thoughtful, transparent, correcting—these are all adjectives used to describe science. Research in psychology, neuroscience, medicine, and other disciplines have been said to be in "crisis" of late (Pashler and Harris, 2012), where the word crisis is used to denote that some of the fundamental assumptions we often (tacitly) make about the scientific literature may not be true. False findings abound (Ioannidis, 2012; Ioannidis and Trikalinos, 2005). More instances of conscious fraud have been exposed (e.g., Bhattacharjee, 2013; Bohannon, 2015); efforts to replicate studies have demonstrated that many findings have not yet been replicated (Open Science Collaboration, 2015); and major journals published studies making improbable claims (Bem, 2011; Wagenmakers et al., 2011). This article aims to briefly review some of the key issues that cast doubt about the conclusions drawn in the psychophysiology literature, with a particular focus on the electrophysiology literature, as well as review potential methods for improving the quality of the literature. My comments are from the prospective of a methodologist who collaborates with researchers in psychophysiology as well as other sub-disciplines in psychology. I also consider how the incentives in academic culture create a context in which change to publishing practices is difficult.

1.1. What counts in academics

Considerations about improving the rigor of research cannot ignore the culture of academia, specifically how that culture shapes the day-to-day activities of scientists. Those activities may seem a bit bizarre to the casual observer, and, if not bizarre, at least unexpected. When I started college my image of a scientist was someone working in a lab, recording data, and running experiment after experiment trying to uncover a truth about the world. I was not aware that scientists often have a team of people that collect and analyze data and that scientists spend much of their time writing papers and grants. Further, I was not aware that one's standing as a scientist largely comes from the number of papers one produces, how much attention those papers engender, and how much money one brings to the university.

In the book *Laboratory Life: The Construction of Scientific Facts*, Latour and Woolgar (1986) document a two-year anthropological study of scientists. Latour spent two years at the Salk Institute in the late 1970s, observing the work of scientists as would an anthropologist observing a previously unknown tribe. Latour strove to take an "outsider's" view of the work, aiming to make sense of varied behaviors, rituals, and norms he observed. After observing the day-to-day work of the scientists, Latour and Woolgar (1986) write:

> The production of papers is acknowledged by participants as the main objective of their activity. The realisation of this objective necessitates a chain of writing operations from a result first scribbled

* Corresponding author at: Brigham Young University, 268 TLRB, Provo, UT 84602, United States.
E-mail address: scott_baldwin@byu.edu.

http://dx.doi.org/10.1016/j.ijpsycho.2016.04.006
0167-8760/

> on a sheet of paper and enthusiastically communicated to colleagues, to the final registering of published literature in the laboratory archives. The many intermediary stages (such as talks with slides, circulation of preprints, and so on) all concern literary production of one kind or another. (p. 71)

After noting the extensive time and resources that are involved in producing these papers, Latour and Woolgar wonder: "…how can a paper be both so expensive to produce and yet so highly valued? What exactly can justify participants' faith in the importance of the papers' contents?" (p. 71–72).

Scientists are judged by their vita and adding publications, especially publications in prestigious journals, is how careers are made. Consider the advice given to new psychology graduate students about the hiring process in *The Compleat Academic*, a popular book aimed at mentoring graduate students and new faculty entering academia:

> The information we need to arrive at a short list of applicants is contained in the letters of recommendation and, primarily, in the academic vita. Wise graduate students, therefore, will start at day one of their first year in a PhD program to develop a strong vita….Alter your perspective so that you derive your professional self-respect entirely from what is on that document. From the start of graduate school on, throughout what we hope will be a long and productive career, you *are* your vita. (Lord, 2004, p. 10, emphasis in original)

I suspect most academics reading this quote would nod in general agreement. To be sure most psychologists want to learn about the world and human behavior. However, the realities of hiring, tenure, grants, and awards, where lines on a vita are paramount to success, are strong and influential. We may not like being reduced to a vita, but, like the scientists observed by Latour and Woolgar, we spend a lot of time producing papers and our careers are judged by those papers.

Publishing papers is not a problem; articles and books are the primary method for scientific communication. Nor is seeking accolades by definition a problem. Science is not a zero-sum game where one is either pursuing knowledge or accolades. However, when one's job is on the line or one is aiming to be the first to publish in a particular area, producing reliable, replicable knowledge may not be as important as producing publishable knowledge. Likewise, when pursuing grant money, questions that are fundable may take priority over questions that are most theoretically relevant.

1.2. Aims

Below I discuss the consequences of researcher flexibility and underpowered studies on the quality of research findings. Although these issues, and the recommendations for how to address them have discussed a lot recently, none is new. We have known about the problems of power, reliance on *p*-values, and excessive researcher flexibility for a long time (e.g., Cohen, 1962, 1994). However, research practices have not changed; we rely on *p*-values as much as ever and we continue to publish underpowered studies. I suspect that some of this "cultural inertia" regarding research practices is associated with how we define success and prestige in academics. Most suggestions for improving the rigor of research lead to fewer publications, more null findings, and more transparency regarding research practices. These reduce the size of our vita, and thus, according to *The Compleat Academic*, our identity as researchers.

The primary aim of this paper is to review some of the latest trends in the methodological literature regarding (a) roadblocks to rigorous research and (b) strategies for improving rigor. Specifically, I discuss (a) researcher flexibility and its impact on the stability of conclusions and (b) the role power plays in the probability that a finding is true and the precision of estimates. I conclude with a discussion of suggestions for addressing these problems. I discuss these recommendations in the context of academic culture because change is not likely to happen without consideration of the day-to-day context in which research occurs.

2. Does psychophysiology need to improve?

I noted previously that the problems regarding replication, rigor, and fraud have been called a "crisis". One might argue that the attention these problems receive is overblown or, at the very least, the problems are limited largely to social psychology and to some extent fMRI. After all, the focus of the Replication Project was social and cognitive psychology (Open Science Collaboration, 2015) and only one study was clearly within the psychophysiology area (Hajcak and Foti, 2008).[1] Psychophysiology may not be similar to these sub-disciplines and may have more robust findings.

This reasoning is problematic because it assumes that psychophysiology is an exception to the problems common to social psychology and other sub-disciplines. Psychophysiology and neuroscience students do not typically receive more methodological training during graduate school than students in other areas. A survey of the top 50 *U.S. News and World Report* psychology programs with a neuroscience degree (36 responded to the survey), showed that neuroscience students are required to take fewer methodology classes than social/personality students (Schwartz et al., 2016). Measurement in psychophysiology is not particularly rigorous. Although psychophysiology measures often are seen as hard evidence because they are measures of physical phenomena, these measures are not particularly strong and sometimes struggle to meet standards of reliability. For example, in fMRI, average reliability is approximately 0.5, averaged across a number of measures of reliability (e.g., test–retest, voxel counts) and tasks (Bennett and Miller, 2010). Two studies have shown that in EEG studies of error-related negativity (ERN), the number of trials needed to obtain reliable estimates exceeds what is often used (Baldwin et al., 2015; Larson et al., 2010). Although there is some disagreement regarding the number of trials needed in ERN research (Foti et al., 2013; Meyer et al., 2013; Olvet and Hajcak, 2009; Pontifex et al., 2010), what is clear is that reliability of these EEG measures is not firmly established.

As I discuss below, statistical power is major issue for the replicability and quality of research findings and low power plagues studies across psychology (Button et al., 2013; Cohen, 1962; Open Science Collaboration, 2015). I am unaware of any evidence that suggests psychophysiology as a whole is well-powered and I present some evidence that suggests just the opposite.

Perhaps psychophysiology research is notably replicable? This is an empirical question and there is no clear evidence to answer one way or another. The replicability project included one psychophysiology study (Hajcak and Foti, 2008). A key finding reported in this paper was that the magnitude of the ERN was negatively correlated with startle responses—making errors is associated with increased startle (Hajcak and Foti, 2008). This correlation was not significant in the Replication Project (https://osf.io/jret9/) nor have the original authors replicated the finding (Riesel et al., 2013). Finally a re-analysis of the data from Hajcak and Foti (2008) indicated that the significant findings disappeared after excluding a single outlier (Moser et al., 2014). This was just one study and one finding. Perhaps this would be unique in psychophysiology research. However, without clear evidence that psychophysiology is particularly rigorous with respect to the design and analysis features that lead to high probability of replication and impact, it is reasonable at this point in time to encourage psychophysiology researchers to increase rigor.

[1] Scholars have debated the merits of the Replication Project (Anderson et al., 2016; Gilbert et al., 2016). Although the Replication Project is not without its problems, it is one part of the broader examination of the problems in psychological research discussed above. Consequently, one does not need to rely on the results of the Replication Project to make the case that much psychological research could be more rigorous.