Marketing Letters (2023) 34:375–382
https://doi.org/10.1007/s11002-023-09679-2

Are scientific practices improving in consumer research? A glass half-full and half-empty

Michel Tuan Pham[1]

Accepted: 19 April 2023 / Published online: 6 May 2023

Abstract
In this commentary, I propose a "big-picture" view of what good science is as a framework for evaluating scientific practices in consumer research. A big-picture view of science recognizes that scientific practices are not ends in themselves but tools to be used in the service of six epistemic ideals: veridicality, precision, transparency, generalizability, relevance, and insight. It is through a multidimensional contribution to these epistemic ideals that various scientific practices enable the production of evidence that is not only trustworthy but also useful. From this big-picture perspective, Krefeld-Schwalb and Scheibehenne's results provide a decidedly mixed report card about the state of scientific practices in consumer research.

Keywords Consumer research · Scientific practices · Open Science · Epistemology · Online research · Mechanical Turk

1 Introduction

As Krefeld-Schwalb and Scheibehenne (2023; hereafter, KSS) noted in their article, a little over 10 years ago the field of consumer research was the subject of fervent calls for reform. Two parallel sets of calls were being issued at the time. One set—the one that KSS's study focused on—revolved around the need to increase the transparency and reproducibility of consumer research (Bazerman, 2012). This first set of calls was issued when it became apparent that the previously unrecognized but possibly widespread use of various questionable research practices (QRPs; John et al., 2012)—for instance, the self-serving construction of dependent measures to attain statistical "significance" or the selective exclusion of data points based on post hoc criteria—likely inflated the rate of false positive results in consumer research (Simmons et al., 2011). As a result of these calls, L.J. Shrum, Connie Pechmann,

✉ Michel Tuan Pham
tdp4@columbia.edu

[1] Business in Marketing at Columbia Business School, Columbia University, 665 West 130th Street, 922 Kravis Hall, New York, NY 10027, USA

and I, along with others, worked on updating the *Journal of Consumer Psychology*'s editorial practices in 2013–2014 to promote greater methodological transparency (Pechmann, 2014; Pham, 2012). Several years later, the *Journal of Consumer Research* adopted similar policies (Inman et al., 2018).

A different set of calls focused on the need to increase the substantive relevance and impact of consumer research (Inman, 2012; Pham, 2013). These calls were motivated by recurring observations that consumer research articles often lack substantive and managerial relevance and typically have very limited impact (see Jedidi et al., 2021 and Pham et al., 2022 for recent results). In my 2013 address to the Society for Consumer Psychology (Pham, 2013), I urged consumer scholars to confront this persistent lack of relevance and impact by avoiding seven common "sins" in consumer research, such as embracing too narrow a definition of consumer behavior and overly focusing on mental processes as opposed to the actual content of consumers' mental processes. Although this second set of concerns is not explicitly examined by KSS's study, some of their findings do speak to these concerns as well. In the present commentary, I outline a "big-picture" view of what good science is as a framework for evaluating scientific practices within our field. I then discuss how KSS's results deliver a mixed report card about the state of scientific practices in consumer research.

2 Toward a holistic assessment of scientific practices in consumer research

As an early supporter of "Open Science" efforts in consumer research (Pham, 2012; see Pechmann, 2014), I do believe that issues of transparency and reproducibility are critical for our field. There is no question that for consumer research to progress as a scholarly discipline, it must embrace strong scientific practices, which include concerns for transparency and reproducibility. However, unlike other advocates of Open Science, I do not believe that issues of transparency and reproducibility can be examined independently of other epistemic ideals such as research relevance and usefulness (Pham & Oh, 2021a, 2021b).

As illustrated by the figure below, adapted from Pham and Oh (2021b), a "big-picture" view of what "good science" is recognizes that it involves a broad range of *scientific tools*, including keen observation, clear hypotheses, skillful experimental designs, rigorous analyses, careful peer-reviewing, and various Open Science practices such as preregistration, replications, and data sharing. These tools are used in the pursuit of *epistemic ideals* such as transparency, veridicality, precision, and generalizability, but also relevance and insight. To the extent that the combined use of various scientific tools enables the fulfillment of these epistemic ideals, the research produces *scientific outputs* in the form of trustworthy evidence. Not all trustworthy evidence is necessarily useful, however; as depicted in the figure, only a subset is *both* trustworthy *and* useful. Trustworthy evidence that is useful then makes *scientific contributions* in the form of knowledge advances, practical applications, and informed inputs to policy decisions and other impacts on society (Fig. 1).

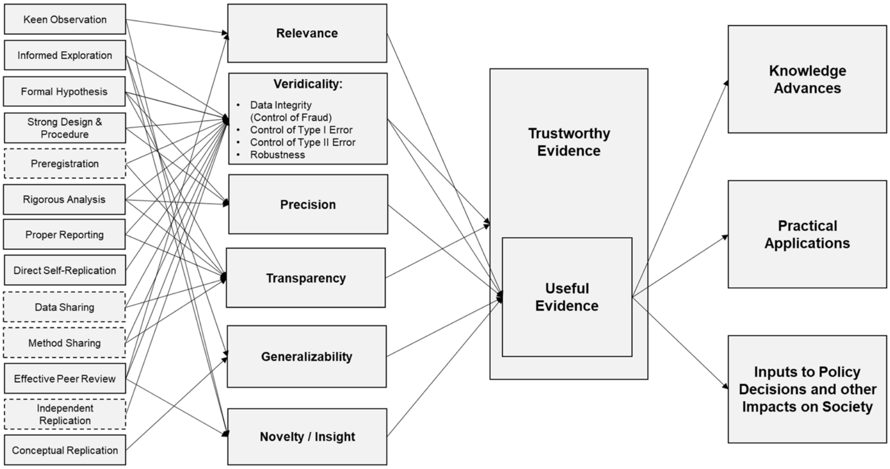

Fig. 1 The "big picture" of good science. (Dashed boxes denote standard Open Science tools.)

This big-picture view of science provides a useful framework for evaluating scientific practices in a field like consumer research. This broader perspective makes it clear that common scientific practices such as random assignment, formal hypotheses, rigorous data analyses, and replications—including newer Open Science practices such as preregistration and data sharing—are just *tools* to be used in the service of higher-order epistemic goals such as veridicality, transparency, and generalizability. None of these tools, including those that fall under the "Open Science" label, should be regarded as an end in itself because no one scientific practice guarantees the production of worthy science. One should therefore beware of interpreting the use of any such tools mechanically as *prima facie* evidence of scientific merit. For instance, not all studies that include a fancy Preacher and Hayes (2008) mediation analysis are necessarily to be trusted, and not all studies that do not include such an analysis are necessarily to be doubted. For the same reason, one should not assume that any study that is preregistered is necessarily "good" or, conversely, that any study that wasn't preregistered is inherently "bad" (Pham & Oh, 2021a).

In addition, a big-picture view of science should remind us that transparency and reproducibility—the main concerns of today's Open Science movement (Nosek et al., 2015)—are not the only epistemic criteria that matter for good science. Robust results that are produced in a transparent manner are not "good science" if they merely demonstrate things that are obvious or substantively irrelevant. Good scientists should not focus only on the transparency and reproducibility of their findings: they should additionally aim at addressing relevant issues, strive to provide novel insights on these issues, be precise in their findings, and establish the generalizability of these findings. Only then will the scientists' evidence be not just trustworthy, but indeed useful.

The recognition that good science is not just about transparency and reproducibility, but also about relevance, precision, generalizability, and insight, brings to light the fact that certain scientific practices may introduce *tradeoffs* among epistemic ideals. For example, direct replications increase confidence in the veridicality of findings but are ill-suited for the demonstration of generalizability. Similarly, flexibility in data analysis may improve the insight potential of a given set of data but do so at the cost of a greater chance of type-I error. Conversely, conservative reviewing policies may reduce the risk of type-I error but do so at the cost of an increased risk of type-II error. Therefore, one should evaluate scientific practices not solely on the benefits they provide along specific epistemic dimensions (e.g., reproducibility) but also based on the potential costs they introduce along other epistemic dimensions (e.g., relevance) (Pham & Oh, 2021b). For example, if, for the sake of transparency and reproducibility, a strong emphasis on preregistration encourages researchers to test only hypotheses that are trivial or blatantly true, science would not be well served (Pham & Oh, 2021a). This broader perspective on worthy science provides a more holistic lens on what KSS's findings reveal about the state of scientific practices in consumer research.

3 Krefeld-Schwalb and Scheibehenne (2023): good news and bad news

In their study, KSS examined the evolution of consumer research practices, from 2008 to 2020, by recording the sample sizes, effect sizes, and distribution of *p*-values of 258 consumer research articles sampled from three major journals during this period. The representativeness of this sample of articles, which were hand-coded, is corroborated by a comparison with an automated text analysis of all articles published in the same journals between 2011 and 2018. I will dispense with minor quibbles about their methodology to concentrate on the big-picture takeaway from their results. Five results attract my attention.

The first is the distribution of *p*-values that KSS observed across the 3,947 hypothesis tests reported by the 258 hand-coded articles (KSS, Fig. 6). One can see that the distribution of these *p*-values is markedly right-skewed, with many more reported *p*-values between 0 and 0.01 (i.e., far below the standard threshold of significance of $\alpha < 0.05$) than between 0.04 and 0.05 (right below the standard threshold of significance). Similar right-skewed distributions were observed for every individual year between 2008 and 2020 (KSS, Web Appendix B). According to "*p*-curve" diagnostic principles (Simonsohn et al., 2014), such a right-skewed distribution of reported *p*-values suggests that, *on average*, there *is* positive evidence against the null hypothesis in consumer research published between 2008 and 2020. Had there been zero evidence against the null hypothesis on average, the distribution of *p*-values would have been flat, whereas had the bulk of the reported findings been driven by extensive use of QRPs (also known as "*p*-hacking"), the distribution would have been *left*-skewed instead (Simonsohn et al., 2014).[1]

[1] As a caveat, it should be noted that the *p*-curve methodology has been criticized by McShane, Bockenholt, and Hansen (2016), who argue that the methodology is unreliable when the sample of studies is heterogeneous, as is the case in the KSS investigation. Simonsohn et al. (2014) maintain that *p*-curve analyses *are* valid even when samples are heterogeneous. However, this disagreement does materially qualify the general interpretation provided here.

Overall, this is good news. With respect to the veridicality of consumer research, it does not seem that the consumer literature *mostly* consists of *p*-hacked, false-positive results, as some may fear based on recent replication failures.[2] However, the suggested interpretation that *most* published consumer research findings are probably not false positives does not mean that there are *no* false-positive results in the consumer literature and that QRPs and *p*-hacking are nonexistent. Indeed, KSS find some evidence of *p*-hacking for a small but non-negligible proportion of the statistical tests performed. For example, a substantial number of tests reported as one-tailed should probably have been performed as two-tailed tests, which likely contributed to some false-positive results.

A second noteworthy finding of KSS's study is a significant increase in the proportion of consumer research articles that include field studies (KSS, Fig. 3). Between 2013 and 2020, the proportion of articles that included at least one field study more than tripled, from less than 10% in 2013 to more than 30% from 2017 to 2020. To the extent that consumer research has often been criticized as being overly dependent on artificial lab experiments (Ferber, 1977; Pham, 2013; Wells, 1993), the increased prevalence of field studies in consumer research articles is a welcome development. Field studies are a useful complement to more controlled experiments. By transcending the narrow confines of the lab, such studies help enhance our confidence in the ecological validity and generalizability of the findings. In addition, because field studies are necessarily grounded in a particular substantive reality, their findings are more likely to have at least some degree of relevance. That said, one would not want *all* consumer studies to be field studies, as such studies typically entail compromises in terms of experimental control, process evidence, causal identification, context specificity, and the like. Again, the desirability of any scientific practice should be judged along multiple epistemic dimensions.

A third notable finding of KSS's study is a substantial increase in sample sizes over the period examined (KSS, Fig. 1). Between 2008 and 2020, the median sample size per condition more than doubled. Although this increase is driven in part by a more frequent reliance on online samples—an issue discussed below—the increase was observed for *both* studies conducted online and studies not conducted online (see left panel of KSS, Fig. 1). Therefore, it appears that in recent years consumer researchers have strived to increase the statistical power of their studies. This too is generally good news. Historically, consumer studies tended to be underpowered, mostly because of the experimental costs of conducting larger studies. While a lack of power is known to increase the chance of failing to observe a true effect in a given study (type-II error), it should be recognized that at the aggregate level, a widespread lack of power across studies increases the odds of false-positive results being published (Ioannidis, 2005). The pervasive lack of power of many past consumer studies most likely contributed to a certain percentage of published results being false positives. It is therefore reassuring that in recent practices, consumer researchers are trying to cast "tighter nets" (to use KSS's expression) for their empirical studies.

[2] See, for example, https://openmkt.org/research/replications-of-marketing-studies/

Here comes the bad news. A key finding of KKS's study is a steady decrease in the effect sizes of consumer research findings over time (KSS, Fig. 2, top and bottom panels). From 2008 to 2020, the median reported effect size in the hand-coded sample dropped from an R^2 of 6.3%—a "medium" effect in line with earlier estimates of typical effect sizes in consumer research (Peterson et al., 1985)—to an R^2 of 2.6%, which would be considered a "small" effect by common standards. A similar reduction was found in the larger text-analyzed sample, with a median effect size dropping from $R^2 = 6.3\%$ in 2011–2013 to $R^2 = 3.6\%$ in 2016–2018. The drastic reduction in the proportion of variance explained by the studies over the past decade suggests that, as KSS aptly put it, while consumer researchers have been casting "tighter nets" in recent years, they simultaneously have been going for "smaller fishes." Therefore, although there have been *efforts* to increase the statistical power of consumer studies, as noted above, these efforts have been largely offset by the smaller size of the effects being studied.

Part of the reduction in effect sizes seems to be attributable to the greater reliance on online studies, which, as KSS observe, are associated with smaller effect sizes ($R^2 = 4.4\%$ for online studies vs. 7.3% for studies not online). However, the drastic reduction in effect sizes over time may reflect substantive changes in research practices beyond the increased reliance on online samples. First, it could be that consumer researchers are indeed pursuing smaller effects, whether willingly or unwillingly, as suggested by KSS's "tighter nets for smaller fishes" title. From an epistemic perspective that places value on relevance and generalizability, this is not good news. Alternatively, it could be that previous estimates of effect sizes in consumer research (which were in the range of $R^2 = 5–6\%$ prior to the 2012–2013 calls for reform) were inflated *all along*. The true effect sizes may in fact be closer to $R^2 = 2–3\%$ after correcting for QRPs, which were presumably curtailed by recent changes in scientific practices (e.g., preregistration, sharing of the original measures, and pre-study power analyses). The real explanation is probably a combination of the two interpretations.

One of KSS's ancillary findings raises another major concern about the evolution of consumer research over the past decade. A major change in the field's scientific practices is a dramatic increase in the proportion of studies that are conducted online as opposed to not online (KSS, Fig. 1, left panel & Fig. 2, top panel). Whereas in 2008–2010, only 1% of the studies was conducted online, by 2018–2020, 50% of the studies were conducted online, mostly through low-cost platforms such as Amazon Mechanical Turk and, more recently, Prolific. Such online studies raise a variety of data-quality issues (see Goodman & Paolacci, 2017; Peer et al., 2022), including the already-noted fact that they tend to be associated with lower effect sizes. From an epistemic perspective, these data-quality issues raise obvious concerns in terms of the veridicality, precision, and generalizability of the resulting findings.

Equally concerning, if not more, is the possibility that the low cost and convenience of online platforms such as MTurk and Prolific is gradually shifting consumer researchers' agendas toward studies that *can* be conducted online, as opposed to studies that ideally *should* be conducted in the first place in order to produce genuine advances in consumer behavior. This a concern that I already raised 10 years ago when I warned the field against the sin of "research by convenience" (Pham, 2013;

see also Ferber, 1977). What happened to consumption phenomena, such as emotion-rich and sensory phenomena, or deep motivational issues, that *cannot* be properly studied on MTurk or Prolific? Are they not worthy of studying anymore? If an omniscient being were to make a comprehensive list of the most pressing and relevant questions to be answered about consumer behavior, would online studies on MTurk or Prolific be the ideal method for as much as 50% of these questions? This is doubtful. Therefore, I genuinely worry that, due to the convenience of online studies, in recent years consumer researchers have not just been chasing smaller effects; they have also been investigating research questions that are less relevant and conducting studies with lower potential for insight.

4 Conclusion

Scientific practices should be evaluated from a big-picture perspective that embraces multiple epistemic criteria and recognizes potential tradeoffs across epistemic criteria. From this broader perspective, KSS's results offer a decidedly mixed report card on the evolution of scientific practices in consumer research over the past decade. On the positive side, their results suggest that although there is evidence of *p*-hacking in consumer research, most published consumer findings are probably not false positives. There is also encouraging evidence that consumer researchers are at least *trying* to increase the statistical power of their studies by recruiting larger samples, and that consumer researchers are more likely to include field studies as part of their empirical packages. On the other hand, it is worrisome that consumer researchers seem to be focusing on effects that are increasingly small. In addition, the recent emergence of online studies as the dominant mode of data collection raises concerns about the overall quality of the data on which consumer research findings are now based, and concerns that researchers' agendas are increasingly being dictated by mere research convenience rather than by a genuine preoccupation with research relevance and insight.

Declarations

Ethical approval Not applicable.

Competing interest The author declares no competing interests.

References

Bazerman, M. (2012). Keynote address. *Society for Consumer Psychology Conference, Las Vegas, February, 18*, 2012.

Ferber, R. (1977). Research by convenience. *Journal of Consumer Research, 4*(1), 57–58.

Goodman, J. K., & Paolacci, G. (2017). Crowdsourcing consumer research. *Journal of Consumer Research, 44*(1), 196–210.

Inman, J. J. (2012). The elephant not in the room: the need for useful, actionable insights in behavioral research. In Z. Gürhan-Canli, C. Otnes, & R. J. Zhu (Eds.), *NA - Advances in Consumer Research, 40* (pp. 1–4). Association for Consumer Research.

Inman, J., Campbell, M. C., Kirmani, A., & Price, L. L. (2018). Our vision for the Journal of Consumer Research: it's all about the consumer. *Journal of Consumer Research, 44*(5), 955–959. https://doi.org/10.1093/jcr/ucx123

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124.

Jedidi, K., Schmitt, B. H., Sliman, M. B., & Li, Y. (2021). R2M index 1.0: Assessing the practical relevance of academic marketing articles. *Journal of Marketing, 85*(5), 22–41.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524–532.

Krefeld-Schwalb, A., & Scheibehenne, B. (2023). Tighter nets for smaller fishes? Mapping the development of statistical practices in consumer research between 2008 and 2020. *Marketing Letters*. https://doi.org/10.1007/s11002-022-09662-3

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science, 11*(5), 730–749.

Nosek, B. A., et al. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425.

Pechmann, C. C. (2014). Editorial regarding the new submission guidelines at the Journal of Consumer Psychology. *Journal of Consumer Psychology, 24*(1), 1–3. https://doi.org/10.1016/j.jcps.2013.10.002

Peer, E., Rothschild, D., Andrew, G., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods, 54*, 1643–1662.

Peterson, R. A., Albaum, G., & Beltramini, R. F. (1985). A meta-analysis of effect sizes in consumer behavior experiments. *Journal of Consumer Research, 12*(1), 97–103.

Pham, M. T. (2013). The seven sins of consumer psychology. *Journal of Consumer Psychology, 23*(4), 411–423.

Pham, Michel Tuan (2012), email to the Membership of the society for consumer psychology announcing the appointment of a taskforce on scientific practices, July 2012.

Pham, M. T., & Travis Tae, O. (2021a). Preregistration is neither sufficient nor necessary for good science. *Journal of Consumer Psychology, 31*(1), 163–176.

Pham, M. T., & Travis Tae, O. (2021b). On not confusing the tree of trustworthy statistics with the greater forest of good science: A comment on Simmons et al.'s perspective on pre-registration. *Journal of Consumer Psychology, 31*(1), 181–185.

Pham, Michel Tuan, Yinghao Alisa Wu, and Danqi Wang (2022), "Benchmarking consumer scholarship: a comparison of alternative metrics," conditionally accepted at Journal of Consumer Research.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891.

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*(11), 1359–1366.

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.

Wells, W. D. (1993). Discovery-oriented consumer research. *Journal of Consumer Research, 19*(4), 489–504.