

**ONCE MORE UNTO THE BREACH:
INTEGRATING REPLICATION INTO RESEARCH PRACTICE**

Oleg Urminsky

oleg.urminsky@chicagobooth.edu

University of Chicago Booth School of Business, Chicago, IL, 60615

First version: 5/2022

*****PLEASE CONTACT AUTHOR FOR THE
CURRENT VERSION BEFORE CITING OR CIRCULATING*****

ABSTRACT

This paper defines the problems of external validity that have motivated calls for more replications, and initiatives to foster replication, throughout the history of JCR. I review the ways in which replication has been featured in consumer behavior and the challenges limiting the extent of replication, using JCR as a specific setting. In particular, standards for external validity and generalizability have increased over time, resulting in growing use of self-replication in original empirical research papers. Independent replication, however, has remained rare. I discuss a constructive framework for how replication, in different forms, could be more fully integrated into different phases of the scientific research process, taking into account contribution necessary for publication. In particular, I identify opportunities to incorporate independent replication into original papers, to increase the replication-based contribution in papers that build on prior work and to use systematic replication in conjunction with meta-analysis to synthesize and confirm conclusions from a mature research literature. More fully integrating replication into scientific practice can yield a new equilibrium, in which replication is routine, typically successful, and an accepted prerequisite for establishing an empirical generalization.

Keywords: External Validity, Generalizability, Replication, Research Contribution, Meta-analysis

“The ability to systematically replicate research findings is a fundamental feature of the scientific process. Indeed, the idea that observations can be recreated and verified by independent sources is usually seen as a bright line of demarcation that separates science from non-science.” (Dunlap, 1926)

A scientific finding is replicable if new relevant empirical evidence consistently yields a comparable result, consistent with the theory to which the evidence applies. Replicability is a promise made by the researcher to the reader, as well as a key criterion for transforming data into facts that can serve as a foundation upon which scientific knowledge is built, and an ongoing concern for any science that aims to be valid and self-correcting. Replicability is also necessary for practical relevance: reliable advice, applications and policies cannot be developed without knowing which findings consistently replicate, under which conditions.

Consider breakthroughs in biomedical research in the 1990s, concerning the role of RNA (Gorman & Moran 2014). If those findings had not consistently replicated, research on the theoretical foundations for medical use of RNA could not have been built on that foundation. If, in turn, that research had not replicated, a new class of RNA-based vaccines to combat the Covid-19 pandemic could not have been developed. Had the promising initial effects of those vaccines not been validated in large-scale clinical trials, we would not be able to walk into a drugstore and get vaccinated, lowering our Covid risk.

Replication is a necessity precisely because it is never a foregone conclusion, as illustrated by the seeming discovery of cold fusion, around the same time. Heralded as a potential breakthrough discovery, cold fusion could not be independently replicated by other labs. As a result, any new research on the topic now starts from the understanding that, even if cold fusion were possible, no replicable method has been discovered (Gibney 2019). Because replication cannot be assumed, a majority of highly cited papers in medicine are the subject of future replication tests (Ioannidis 2005). After all, the more useful a finding is to its field, the

more important it is to be sure of its validity, and the more value there should be in testing replication, in some form. However, in many fields, including consumer behavior, replication is not necessarily a well-integrated aspect of the process of scientific development.

Discussion of replication is increasingly common in consumer behavior, and more specifically in the *Journal of Consumer Research* (JCR). Compared to only a quarter of papers in JCR's first 10 years, 65% of papers in the last ten years mentioned replication. Despite the centrality of replication to consumer behavior research, there has also long been dissatisfaction with the extent of replication in the field and repeated attempts to encourage more replication.

Next, I define the problems that replication is intended to address and review the ways in which replication has featured in consumer behavior and the challenges limiting the extent of replication, using JCR as a specific setting. I then provide a constructive framework for how replication, in different forms, could be more fully integrated into phases of the scientific research process, taking into account publication standards for sufficient contribution.

THE NEED FOR REPLICATION

The Problems Underlying the Need for Replication

I begin by defining the problems replication is intended to remedy. Consider the testable implication of a theory: differences in X should cause (or predict) a change in Y :

$$Y = a + b \cdot X + e \quad (1)$$

This model is empirically testable by measuring X and Y or by manipulating X and observing the resulting value of Y . The inclusion of the error term e means that we cannot expect a change in X to result in the same change in Y every time. Instead, b represents an *average estimated* effect, and not “the” effect of X on Y . Finally, it is important to note that treating e as normally distributed *random* error is an assumption of convenience. By definition,

e represents deviation of actual values of Y from the corresponding predicted value ($a + b \cdot X$), which includes *all omitted other causes* of Y that may vary in the world.

In fact, the estimate of b from a given data set may not only vary randomly due to sampling error but is highly likely to also vary systematically. Thus, we should think of b as not only a noisy average, but as a construct that can depend on multiple other factors:

$$b = b_0 + b_P \cdot P + b_S \cdot S + b_C \cdot C + b_M \cdot M + b_T \cdot T + b_R \cdot R + e \quad (2)$$

This formulation represents a hierarchical linear model, in which the estimated coefficient of interest is decomposed into the various factors that cause (or predict) variation in the coefficient (Yarkoni 2022). Each b_k may be a vector of coefficients, corresponding to a set of predictor variables. Equation 2 specifies common types of variables across which the coefficient may vary: P for sample population characteristics, S for stimuli, C for contextual factors, M for measures used in the research, T for time-varying factors and R for researcher-specific factors.

The fundamental problem motivating replication is that a test will yield a *specific* estimate of b that depends on the values of P, S, C, M, T & R in that test, but the researcher may mistakenly treat it as an *average* estimated effect (or, even worse, as the *singular* effect of X on Y). When other researchers then apply what they think has been learned to a new theoretical or applied setting (i.e., where P, S, C, M, T , and/or R are different), they will have no idea to what degree the conclusions will still apply. Without an understanding of how b varies with P, S, C, M, T , and/or R , applications of the estimated effect b are based on assumption, not evidence.

Thus, replication is motivated by the twin questions of whether what has been measured is the finding that we think it is and whether it will generalize as assumed. The need for replication arises from a problem of potential heterogeneity, which calls into question the validity and generalizability of a specific finding (Lynch 1982; Monroe 1991; Wells 2001;

LeBel et al 2017; Fabrigar, Wegener & Petty 2020; Stromland 2021). If diverse replications indeed confirm that the parameter of interest is highly homogeneous (e.g., in the extreme, that $b = b_0$), then concerns about generalizability and validity will have been addressed. However, the range of potentially relevant factors, as indicated in (2), highlights the challenge of conducting sufficient replications to provide the evidence needed to form reasonable beliefs about b_k .

If one of the factors in Equation 2 does not vary within or across prior studies, it is impossible to assess the heterogeneity of the effect along that factor, and the representativeness of the estimated purportedly “average” effect b is unknown. Wells (2001) makes an analogy between participant sampling and other experimental factors. Just as we would not assume that a result obtained with a single research participant ($N=1$) applies to people in general, drawing conclusions when $N=1$ on other relevant dimensions (e.g., P, S, C, M, T, and/or R) requires making strong assumptions about homogeneity of the effect along those dimensions.

One set of parameters, b_R , representing the sensitivity of b to researcher-specific factors is particularly relevant to debates about replication. In the scientific ideal, $b_R = 0$, while a nonzero b_R represents a systematic researcher-specific bias. In practice, b_R may be nonzero because the model is mis-specified, failing to include other factors confounded with researcher (e.g., a specific participant population used by that researcher). Even in a properly specified model, however, the true b_R will not always be zero. “Questionable research practices” can bias b_R , intentionally or not, in a researcher-desired direction (Simmons et al 2011, John et al 2012).

Can Replication Address Problems of Potential Heterogeneity and Lack of External Validity?

In theory, replication provides the answer: the more evidence collected while varying P, S, C, M, T, and/or R, the more information we will have about the distribution of realizations of the effect of interest. The need for replication, however, generates a host of practical

questions, including how replications will be conducted, at what stage of the research process, by whom and under what incentives, and whether or where they will be published.

McGrath and Brinberg (1983) divide research into three phases: initial exploration to inform research design, confirmatory execution in which the effect of interest is tested, and systematic exploration of range and limits. Replication fits most readily in that third, often neglected, phase (Wells 1993), and may be of limited usefulness when it occurs before a research area has been sufficiently developed (e.g., when the theory is mis-specified or there is a mismatch between theory and tests; van Rooij & Baggio 2021).

Replications further fall onto a continuum, based on the degree of difference in the research methods (e.g., P, S, C, M, T). Direct replications (i.e., exact, strict, or statistical replications) attempt to keep everything the same, to test whether a prior result was a statistical fluke and to test for researcher-specific factors. At the other extreme, conceptual replications attempt to devise a novel test of the theory, deliberately changing P, S, C, M and/or T relative to the original study. While conceptual replications provide more information regarding the distribution of the effect of interest across the varying factors, when different results are found, the researcher may not know which factors are responsible for the difference. Replications are further distinguished between self-replication (by the same researchers) and independent replication (by different researchers) and between simultaneous (same T) vs. sequential (different T) replications (e.g., Hunter 2001). As these distinctions illustrate, multiple different kinds of replications may be required to investigate the variation in the effect of interest.

Concerns about insufficient replications being conducted in consumer behavior are as old as JCR itself. A 1976 review of consumer research complained that “there is all too little programmatic research and too little replication and cross-validation of findings” (Jacoby

1976). JCR has periodically featured calls for more replications (Helgeson et al 1984; Monroe 1991 & 1992; Hunter 2001; Mick 2003; Rapp 2015), often discussing the disincentives to do or publish replications, urging greater concern about external validity and defining desirable characteristics of publishable replications. Specific initiatives have also been attempted to encourage more replication research in JCR (e.g., establishing a “Research in Brief” section, Bettman & Kassajian 1982; emphasizing editorial support for publishing replications, Monroe 1991 & 1992; establishing a “Re-inquiries” section, Mick 2001), only to be discontinued later.

While the supply problem of independent replications is multiply determined, the criteria used to assess contribution in a highly selective journal like JCR is one key factor. Even those promoting more replication have emphasized the need for sufficient contribution, based on the importance of the phenomenon, the novelty and influence of the paper being replicated, and the rationales provided for potential lack of replication or boundaries to the phenomenon (Monroe 1991 & 1992; Mick 2001). One challenge for addressing external validity via replication is that a substantial proportion of papers may be considered important enough to publish, but not important enough to then publish a replication, whether successful or not.

The goal of external validity can conflict with other criteria for evaluating contribution. For better or worse, novelty, uniqueness, interestingness and surprisingness are often seen as key attributes of high-contribution research. Some guidance even explicitly advises against “exact replication of an earlier study” (Sterthal, 1994), and undervaluation of replication has been attributed to “an overemphasis on creativity” (Hunter 2001; see also Mick 2003). This conflict between novelty and generalizability may explain the seemingly cyclical nature of concern about replication. Selecting on papers that subjectively provide the highest original contribution will shift the literature away from replication, until concerns about replicability

and external validity become sufficiently troubling to re-emphasize the need. However, without replication, the seeming contribution of high-novelty papers may be illusory (Campbell 2017).

How Replication Has Been Discussed in JCR.

I conducted a systematic analysis of all papers published in the Journal of Consumer Research since the founding of the journal.¹ Out of 2636 articles, 1277 (48.4%) had some discussion of replication. As shown in Figure 1, discussions of replication have been increasing dramatically over time ($r=.83, p<.001$).

Figure 2 breaks down the different types of replication discussions in JCR papers. Many papers (38%) discuss whether an empirical result in the paper replicates a finding. However, these discussions are overwhelmingly of self-replications, testing whether the results of one study replicate the results of another study in the same paper (in 32% of papers), and typically reporting a positive replication of their own results. Reporting non-replications has become more common recently (typically in a web appendix) but remains fairly rare.

By contrast, the rate of independent replication is low. Only 5% of papers discuss whether a result in the paper replicates a result by different authors. The majority of these mentions are in papers that provide a test of replication because of some overlap with prior research, but the replication is a primary aspect of the contribution in very few papers. Beyond reporting empirical results, 20% of papers discuss replication more generally (e.g., replications in specific other papers or in general, or calls for future replications of own or others' findings).

As can be seen in Figure 1, the dramatic increase in discussion of replication over time can be traced back to a single source: self-replication. While rare in the early years of JCR, at around 10% of papers, self-replication has now become a norm, with around 60% of papers

¹ Additional details and data are publicly available at https://osf.io/my2d3/?view_only=4ea740de89e9458f8885d7459fa4fdcc

(and an even higher proportion of empirical papers) discussing replications of their own results ($r=.91, p<.001$). By contrast, independent replication has remained similar over time ($r=-.07, p=.63$), as have general discussions of replication ($r=.17, p=.24$). There is no indication that either prior specific initiatives at JCR (e.g., Bettman & Kassajian 1982; Monroe 1992; Mick 2001) or the recent replication crisis in psychology have had a detectable effect on the prevalence of independent replication.

Figure 1: Overall and Specific Changes in JCR Replication Mentions

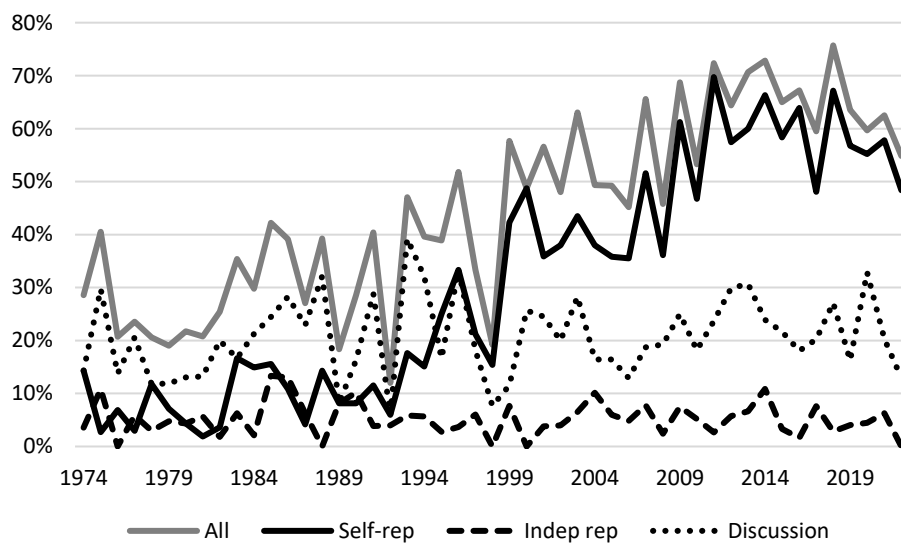
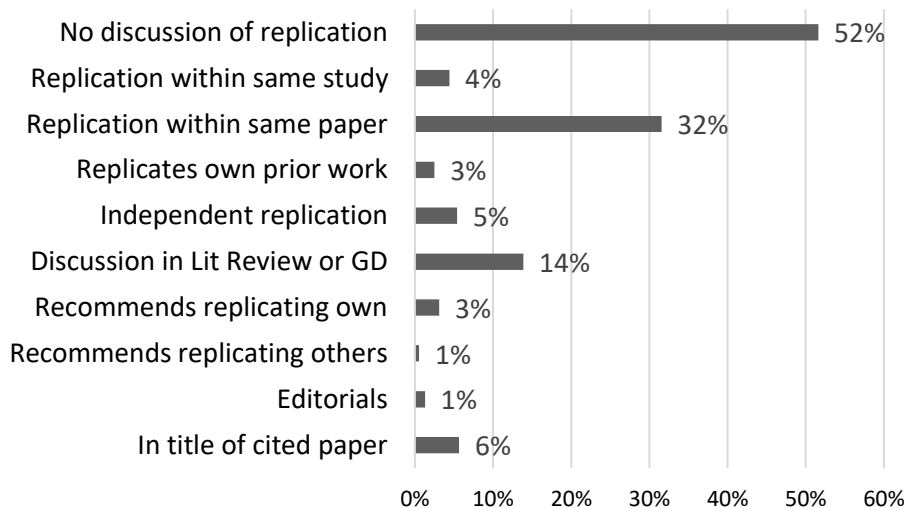


Figure 2: Share of JCR Publications With Each Type of Replication Mention



The dramatic increase in self-replication is, at least in part, a response to prior critiques of how consumer behavior research had been done (e.g., Lynch 1982, Wells 2001). Empirical JCR papers now use more tests than before, varying at least some of the P, S, C, M and T dimensions (e.g., sampling from more than one population, using multiple stimuli or tasks and measures, and testing different contexts at different times). The persistently missing element, however, is variation in R: independent replication is rare, and exact replications or systematic conceptual replications (as opposed to incidental conceptual replications) are rarer still.

This data on the past fifty years of JCR makes it clear that, intentionally or not, the field of consumer behavior has addressed external validity concerns primarily through self-replication within the original paper. Higher standards for publication increasingly dictate that “Stage 3” of research development (e.g., “replication and systematic search for range and limits,” Wells 1993) takes place within the original paper. This approach has advantages: introducing the replication evidence alongside the proposed theoretical advance should be a faster, more standardized and more efficient alternative to relying on independent replication. However, even the substantial investment in time, effort and resources required to incorporate extensive self-replication in original papers is unlikely to solve the external validity problem.

There are three problems with relying primarily on self-replication. First, by definition, self-replications keep R fixed and therefore cannot identify whatever researcher effects may be present. Second, researchers advancing a particular theoretical agenda (and perhaps trying to convince a skeptical review team) are likely, intentionally or not, to skew their choice of research design factors (P, S, C, M, T) towards configurations that favor observing the effect of interest. Sometimes, as when the stated research goal is an “existence proof,” this may be directly discussed in the paper, while in other cases the likely non-representativeness may go

unacknowledged. Third, publication bias can exacerbate these effects, favoring papers with highly consistent results over messier papers with some null findings. As a result, the literature as a whole may be biased towards papers that less thoroughly test generalizability.

Ultimately, whether self-replication is sufficient or not is an empirical question. Multiple multi-site pre-registered replications of highly cited findings in the behavioral sciences have revealed a substantial level of failures to replicate (e.g., about half of findings replicated in Klein et al 2018). Given the overlap in methods and standards with related areas of psychology, consumer behavior research cannot be assumed to have substantially higher replication. Indeed, non-replications have been observed when consumer behavior findings have been the subject of pre-registered replications. For all its benefits, self-replication alone seems to not have fully addressed the validity problems.

TOWARDS AN INTEGRATION OF REPLICATION INTO RESEARCH PRACTICE

It's worth stating explicitly that the preceding discussion is not at all novel -- these points have been made in JCR before, and probably better (e.g., Lynch 1982, Monroe 1991 & 1992, Hunter 2001, Wells 2001, Mick 2003). What this long history suggests is that simply pleading, yet again, for the field to place a higher value on replication is unlikely to make a difference. If replication research remains the metaphorical "dental flossing" of science (*i.e.*, a necessary but tedious and distasteful aspect of hygiene), a repeated cycle of neglect and overconfidence followed by generalization failures and increased concern will result.

Making constructive progress may require rethinking research contribution by flexibly re-situating replication, in different forms, throughout the scientific process. As a starting point, we can think about research contribution in general, and the benefits of replication in particular, in terms of Bayesian updating (Raman 1994): information is valuable to the degree that it leads

to updated beliefs, either regarding the mean (the average effect), the conditional mean (the predicted effect in a given circumstance) or the variance (the uncertainty or heterogeneity around an effect). Crucially, scope of contribution depends on the change in *objective* belief (what *should* be believed based on existing evidence), not changes in subjective beliefs (one's surprise relative to potentially mis-calibrated prior beliefs). From this vantage point, I next discuss how to integrate different forms of replication throughout the scientific process.

Approaches to Integrating Replication Throughout Research Practice.

Self-replication. The increase in self-replication (typically within a single paper) in consumer behavior research should be recognized as a significant accomplishment to be maintained, and a clear indication of the field's concern with external validity. The potential for self-replication to provide the needed "database of facts" (Hunter 2001) could be dramatically increased by making it standard to report (e.g., in an online repository) all studies collected, along with well-documented data and analysis code. For self-replication to be the primary mode of exploring external validity, establishing the comprehensiveness and credibility of self-replication must become a primary goal in the review process, on par with questioning internal validity (confounds in experimental design) and construct/argument validity (theoretical development).

Nevertheless, it is unlikely that self-replication alone will be sufficient. External validity could be more thoroughly established within original papers, but it would likely require new methods and approaches. In much the same way that researchers often rely on independent coders unaware of the hypotheses, papers could include a replication study conducted by an independent replicator, to address researcher effects. Journals could also expand the role of peer review, conducting replication during the review process. Such departures from standard practices would be needed to render post-publication replication unnecessary.

Independent Replication Commentaries. Arguably, sometimes independent replications should be small in scope, empirically and/or theoretically (e.g., relative to the original paper being replicated). While such replications are informative and valuable to the field (particularly when well-conducted, pre-registered and sharing data), it may not be clear where such work will be published. Self-publishing (e.g., blog posts) misses the benefits of peer review, potentially makes the work harder to find (particularly in the long term) and may be more easily dismissed by the field. However, the journal that published the original paper may see the replication as insufficient in scope, while other journals may see the replication as not their responsibility. Arguably, the original journal does shoulder an obligation that may be best fulfilled by publishing brief commentaries or letters (e.g., as done in *Psychological Science*, Bauer 2021).

Replication Within Subsequent Original Research. The scope of a replication project can be expanded via “replicating and extending” -- in effect, combining (near-)exact replication with conceptual replication to test well-motivated boundary conditions or generalizations and applications. However, the degree to which the field should prioritize deeper investigations of existing constructs versus development of new constructs has been a long-running debate (Helgeson et al 1984; Hunter 2001; Shavitt 2011; Campbell 2017). Such approaches risk being seen as “incremental” if the original paper already contains self-replications and moderators.

However, research is cumulative and most empirical papers do build on prior research, both in theoretical development and in the design of stimuli and experiments. As a result, most research contains studies that are, in part, conceptual replications of prior work. JCR papers that include independent replication tend to be framed around a broader or more novel original contribution. A greater appreciation of the factors that make replications particularly valuable, by both authors and reviewers, could improve the contribution of such work.

Studies could be better designed, with comparability to prior research and the resulting replication-based contribution in mind. Replication-relevant data (*e.g.*, studies that attempted to use a manipulation from the prior literature but failed) should be mentioned and shared in an online repository. An accounting of the replication-based implications, both positive and negative, of the paper's findings could be a common feature of the general discussion. These approaches would inform the field about the generality of prior research from new papers.

However, it is important to note the potentially substantial disincentives to authors of taking this approach, primarily the risk of alienating reviewers whose work had not replicated. For this approach to generate a *representative* dataset of independent replications from original research (*vs.* only successful replications) requires reviewers and other stakeholders to both actively encourage authors to develop the replication-based contribution and actively discourage prior authors from suppressing undesired non-replications.

Research Synthesis Through Meta-Analytic Replication. Recognizing that replication is largely an issue of external validity and generalizability suggests that the greatest need for systematic replication may, in fact, come late in the research process. It is rare that a single initial paper provides the last word on any topic. Science is typically cumulative, with general conclusions derived from comparison and synthesis of all the findings in a literature, in the form of qualitative (theoretical reviews) or quantitative (meta-analytic reviews) analyses. However, even a mature research literature may be comprised of either insufficient or systematically biased data points, due to publication bias or disincentives to conduct or publish replications.

This can be remedied by making replication a key element of conducting research synthesis. Simply put, if a field is ready for a meta-analysis on a particular question, the criteria for replication research to make a sufficient contribution are likely to have been met (*i.e.*, the

phenomenon has been deemed important, the prior work is novel and influential, and conflicting results or moderators identified in the prior work make the case that replication is needed.) Pre-registered, sufficiently powered and independently conducted replications should provide exactly the kind of unbiased estimates of the overall effect that typically cannot be assumed from traditional meta-analyses (Kvarven et al 2020). Confirmatory replication should be conducted to confirm the conclusions of meta-analyses (either within the same paper or as a follow-up to meta-analysis) for a field to characterize a finding as an empirical generalization. Meta-analytic replication has been used to form conclusions about ego-depletion (Vohs et al 2021), psychological scarcity (O'Donnell et al 2021) and pain of paying (Bechler et al 2022).

Meta-science Replications. Recent high-profile replication research has largely taken a very different approach from any of the ones outlined above. The “Many Labs” studies, perhaps the most prominent example, conducted multi-site large-scale replications of a selection of (largely unrelated) prominent research findings (*e.g.*, Klein et al 2018). The primary goal was not to deepen our understanding of any particular finding, but instead to use those findings as a basis for exploring meta-scientific questions regarding replication rates and sensitivity to research settings or expertise. While effective at demonstrating the problem of non-replication, it is not a useful model for replication work that aims to improve external validity, because the resources required ensure that only a small subset of studies could be replicated in this way. Perhaps meta-scientific replications should be seen as an indication that the field has failed to foster norms of self-correcting science to effectively address replication and external validity throughout the scientific process in more mundane but scalable ways.

Who Should Conduct Replications?

In an ideal world, all empirical researchers would conduct replications as a standard

practice, in the course of conducting their own research. Since researchers rely on stimuli, tasks, measures, findings and theories from others' work, replication should be done in the design phase or while answering a new question. In due course, the replications would then appear in the research literature, in whichever of the forms discussed above are most fitting.

It is important to recognize the disincentives that can distort what appears in the literature. It is human nature for researchers, particularly junior researchers, to be concerned about offending other researchers in the field, who may be their co-authors, colleagues, reviewers, editors, potential employers of themselves or their students, or tenure evaluators. Paradoxically, when replication is taken seriously, the fact that non-replication may constitute a severe criticism (as opposed to just another data point) or may even be taken as a tacit accusation of fraud, means that it can be problematic to publicize an incidental non-replication finding until one is quite certain. Attempting to publish replications may be uncertain, costly, and frustrating, using scarce resources that a researcher may prefer to prioritize differently. Researcher's reluctance to conduct and publicize replications can therefore run the gamut from reasonable prioritization within one's research agenda all the way to tacit collusion.

These all-too-human factors cut against the effective conduct of science and distort the evidentiary basis upon which we all rely. What can then arise is a specialization of a small set of "replicators," in contrast to "researchers," the majority. Plainly put, this is a symptom of dysfunction in a field that has failed to do self-correcting science, with distinct negative consequences. Unless formally integrated into systematic review processes, "policing" of a field by those willing to conduct replications is likely to result in diffusion of responsibility, dismissiveness of undesired replication results, mutual antagonism and skepticism of motives, and actual or perceived arbitrariness of what is replicated. Ultimately, it is the responsibility of

all researchers, especially those protected by tenure, to maintain research standards and ensure generalizability of knowledge within their own area of study and expertise.

Conclusion.

Replicability enables consumer behavior research, as a cumulative science, to build on prior findings, incorporate substantiated theories and validated methods into new research, and correctly identify important new questions. Replicability is also necessary for having practical relevance to consumers, practitioners and policy makers. Reliable applications cannot be developed and reliable guidance cannot be given without knowing whether a finding consistently replicates, and under which conditions.

The dramatic growth of self-replication, as evidenced by substantial changes in the scope and content of JCR articles over the past 50 years, has improved the external validity of consumer behavior research and facilitated identifying important boundary conditions. However, the relative lack of independent replication has hampered the ability of the field to establish empirical generalizations, contributing to an inability to distinguish between replicable and non-replicable research. To further improve external validity, independent replication should be incorporated throughout the research process, in some combination of original papers, follow-up research that builds on those papers and, perhaps most promising, in meta-replication reviews that synthesize prior research and conduct confirmatory tests of the conclusions. Integrating replication throughout scientific practice can yield a new equilibrium, in which replication is routine, typically successful, and a prerequisite for establishing an empirical generalization.

REFERENCES

- Bauer, Patricia J. (2021) "A New Option for Scientific Exchange and an Alternative to the Commentary Format." *Psychological Science* 32, no. 9: 1343-1345.
- Bechler, Chris, Rhia Catapano, Szu-chi Huang, and Oleg Urminsky (2022) "The Pain of Paying Effect Revisited: Replicating the Past, Moderating the Present, and Informing the Future"
- Bettman, John & Harold Kassirjian (1982) "Editorial," *Journal of Consumer Research*
- Campbell, Meg (2017) "ACR Presidential Address," *ACR Proceedings*, v45 p. 86.
- Dunlap, K. (1926) The experimental methods of psychology. In: Psychologies of 1925, ed. C. Murchison, pp. 331–53. Clark University Press.
- Fabrigar, Leandre R., Duane T. Wegener, and Richard E. Petty. (2020) "A validity-based framework for understanding replication in psychology." *Personality and Social Psychology Review* 24, 4: 316-344.
- Gibney, Elizabeth. (2019) "Google revives controversial cold-fusion experiments." *Nature* 569, no. 7758: 611-612.
- Gorman, Christine & Dina Moran (2014) "How RNA Discoveries Are Radically Changing Gene Therapy and Other Medical Treatments," *Scientific American*, pp. 39 – 45.
- Helgeson, James G., E. Alan Kluge, John Mager, and Cheri Taylor. (1984) "Trends in consumer behavior literature: a content analysis." *Journal of Consumer Research* 10, no. 4: 449-454.
- Hunter, John E. (2001) "The desperate need for replications." *Journal of Consumer Research* 28, no. 1: 149-158.
- Ioannidis, John PA. (2005) "Why most published research findings are false." *PLoS Medicine* 2 (8)
- Jacoby, Jacob. (1976) "Consumer psychology: An octennium." *Annual Review of Psychology* 27, no. 1: 331-358.

- John, Leslie K., George Loewenstein, and Drazen Prelec. (2012) "Measuring the prevalence of questionable research practices with incentives for truth telling." *Psychological Science* 23, no. 5: 524-532.
- Klein, Richard A. et al. (2018) "Many Labs 2: Investigating variation in replicability across samples and settings." *Advances in Methods and Practices in Psychological Science* 1, no. 4: 443-490.
- Kvarven, Amanda, Eirik Strømmland, and Magnus Johannesson. (2020) "Comparing meta-analyses and preregistered multiple-laboratory replication projects." *Nature Human Behaviour* 4, v4: 423-434.
- LeBel, Etienne P., Derek Berger, Lorne Campbell, and Timothy J. Loving. (2017) "Falsifiability is not optional." *Journal of Personality and Social Psychology* 113, no. 2: 254-261.
- Lynch Jr, John G. (1982) "On the external validity of experiments in consumer research." *Journal of Consumer Research* 9, no. 3: 225-239.
- McGrath, Joseph E., and David Brinberg. (1983) "External validity and the research process: A comment on the Calder/Lynch dialogue." *Journal of Consumer Research* 10, no. 1: 115-124.
- Mick, David (2001) "Editorial," *Journal of Consumer Research*
- Mick, David (2003) "Editorial," *Journal of Consumer Research*
- Monroe, Kent (1991) "Editorial: On Replications in Consumer Research, Part I," *Journal of Consumer Research*
- Monroe, Kent (1992) "Editorial: On Replications in Consumer Research, Part I," *Journal of Consumer Research*
- O'Donnell, M., Dev,... & Nelson, L. D. (2021). Empirical audit and review and an assessment of evidentiary value in research on the psychological consequences of scarcity. *Proceedings of*

the National Academy of Sciences, 118(44)

Raman, Kalyan. "Inductive inference and replications: A Bayesian perspective." *Journal of Consumer Research* 20, no. 4 (1994): 633-643.

Rapp, Justine M., and Ronald Paul Hill. (2015) "'Lordy, Lordy, Look Who's 40!'" *The Journal of Consumer Research Reaches a Milestone.* *Journal of Consumer Research* 42, no. 1: 19-29.

Shavitt, Sharon (2011) "What's New? Novelty in Consumer Research" *ACR Proceedings*.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. (2011) "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant." *Psychological Science* 22, no. 11: 1359-1366.

Sternthal, Brian (1994) "Editorial," *Journal of Consumer Research*

Strømmland, Eirik. (2021) "Making our "meta-hypotheses" clear: heterogeneity and the role of direct replications in science." *European Journal for Philosophy of Science* 11, no. 2: 1-19.

van Rooij, Iris, and Giosuè Baggio. (2021) "Theory before the test: How to build high-verisimilitude explanatory theories in psychological science." *Perspectives on Psychological Science* 16, 4: 682-697.

Vohs, Kathleen D., Brandon J. Schmeichel, Sophie Lohmann, Quentin F. Gronau, Anna J. Finley, Sarah E. Ainsworth, Jessica L. Alquist et al. (2021) "A multisite preregistered paradigmatic test of the ego-depletion effect." *Psychological Science* 32, no. 10: 1566-1581.

Wells, William D. (1993) "Discovery-oriented consumer research." *Journal of Consumer Research* 19, no. 4: 489-504.

Wells, William D. (2001) "The perils of N= 1." *Journal of Consumer Research* 28, no. 3: 494-498.

Yarkoni, Tal. (2022) "The generalizability crisis." *Behavioral and Brain Sciences* 45.