ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Hubbard, Raymond

**Working Paper**

# A proposal for replicating Evanschitzky, Baumgarth, Hubbard, and Armstrong's "Replication research's disturbing trend" (Journal of Business Research, 2007)

Economics Discussion Papers, No. 2017-75

**Provided in Cooperation with:**
Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

*Suggested Citation:* Hubbard, Raymond (2017) : A proposal for replicating Evanschitzky, Baumgarth, Hubbard, and Armstrong's "Replication research's disturbing trend" (Journal of Business Research, 2007), Economics Discussion Papers, No. 2017-75, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:
https://hdl.handle.net/10419/169134

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

http://creativecommons.org/licenses/by/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Economics
The Open-Access, Open-Assessment E-Journal

**Discussion Paper**
No. 2017-75 | September 28, 2017 | http://www.economics-ejournal.org/economics/discussionpapers/2017-75

# A proposal for replicating Evanschitzky, Baumgarth, Hubbard, and Armstrong's 'Replication research's disturbing trend' (Journal of Business Research, 2007)

*Raymond Hubbard*

**Abstract**
This paper is about how the author proposes to replicate Evanschitzky, Baumgarth, Hubbard, and Armstrong's 'Replication research's disturbing trend' (*Journal of Business Research*, 2007). This is because estimating the incidence of published replication research and its outcomes must be continued.

(Published in Special Issue The practice of replication)

**JEL** A14 B23 B41 C18
**Keywords** Replication; replication with extension; statistical power analysis; overlapping confidence intervals

**Authors**
*Raymond Hubbard*, ✉ College of Business & Public Administration, Drake University, Des Moines, Iowa, USA, drabbuhyar@aol.com

**Citation** Raymond Hubbard (2017). A proposal for replicating Evanschitzky, Baumgarth, Hubbard, and Armstrong's 'Replication research's disturbing trend' (Journal of Business Research, 2007). Economics Discussion Papers, No 2017-75, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2017-75

**Received** August 23, 2017 **Accepted as Economics Discussion Paper** September 12, 2017
**Published** September 28, 2017
© Author(s) 2017. Licensed under the Creative Commons License - Attribution 4.0 International (CC BY 4.0)

## (i) General Principles for Conducting a Replication

When carrying out an *exact* or *strict* replication, we must first assume that the research in the original article has been done competently. Then, the idea is to duplicate as closely as possible the same procedures used in that article. An exact replication involves taking another sample from the *same* population employed in the earlier work. This definition of an exact replication was offered by Hubbard and Armstrong (1994, p. 236), and is also employed by Tsang and Kwan (1999) in their typology of replications shown in the Table below.

TABLE—A Typology of Replications

| *Source of Data* | *Same Measurement and Analysis* | *Different Measurement and/or Analysis* |
|---|---|---|
| Same data set | (1) Checking of analysis | (2) Reanalysis of the data |
| Same population | (3) Exact replications | (4) Conceptual extensions |
| Different population | (5) Empirical generalizations | (6) Generalizations and extensions |

Source: Tsang & Kwan (1999, p. 766)

In practice, exact replications are never conducted because of the impossibility of duplicating every circumstance. The passage of time alone between the publication of the original article and the appearance of its follow-up ensures this. Or as Ziman (1978, p. 56) notes, "One cannot step twice into the same river." So, by necessity, there will always be *some* differences between the two studies, with the more realistic goal being to minimize them.

Because of the editorial-reviewer bias against publishing replications (see Hubbard, 2016, pp. 158-164 for a review of these), the latter must also incorporate elements of "originality." This is why Hubbard and Armstrong (1994) prefer the expression "replication with extension" to exact replication. The extensions' originality generally comes from modifications to the measurement instrument (enhancing construct validity)—what Tsang and Kwan (1999) call

*conceptual extensions*—and/or the target (sub)population (enhancing empirical generalization) if the replication is "successful."

That said, in this particular case I am in a position to attempt as close to an exact replication of a prior work as is likely to happen. My proposal is to replicate Evanschitzky, Baumgarth, Hubbard, and Armstrong's (2007)—hereafter EBHA—study of the frequency of replication research published in leading marketing journals, which itself is a replication of Hubbard and Armstrong (1994)—hereafter HA. I am aware of the requirement to select an "article that has not previously been replicated" (Reed, 2017, unpaged). My rationale for replicating a previously replicated article is provided below.

### (ii) Why Replicate EBHA?

Replication research is vital to the integrity of science. It is the primary means for assessing the validity, reliability, and generalizability of scientific findings. Yet it is something honored more in the breach than in practice. Incredibly, as mentioned above, there is strong evidence of an editorial-reviewer bias against publishing replications. Consequently, it is important to estimate from time to time how much of such work shows up in the journals of the various disciplines.

EBHA replicated HA's study which examined how often replications were published in three leading marketing journals—*Journal of Consumer Research (JCR), Journal of Marketing (JM),* and the *Journal of Marketing Research (JMR)*—over the 16-year period 1974-1989. They did so based on a content-analysis of 31 randomly selected annual issues from each journal. This 50% sampling of all *JCR, JM,* and *JMR* issues yielded a total of 835 empirical articles. HA found no exact replications in their sample; they found only 20 (2.4%) replications with extensions.

According to the authors of the replications, who typically based their decisions on the outcomes of statistical significance tests, HA further reported that of the 20 extensions, 60% (12) conflicted with the original article, 25% (5) provided some support, and only 15% (3) confirmed earlier results. While almost universal, using *p*-values to determine the success or otherwise of a replication is *not* a recommended procedure.

Over a period of many months, HA attempted to publish their manuscript in the *JCR*, *JM*, and *JMR*. It was rejected by each journal in turn. HA eventually published their work in the *International Journal of Research in Marketing*.

With all four committed to the crucial role(s) of replication research, EBHA decided to update HA's findings. Based on a *census* of all 1,389 empirical articles published in the *JCR*, *JM*, and *JMR* over the 15-year period 1990-2004, they estimated the publication frequency of replications with extensions to be 1.2% (16). In other words, their estimate is only one-half that of HA's lamentable figure.

If any solace can be gleaned from EBHA it is the fact that, for reasons unknown, the outcomes were less depressing than HA's. Of the 16 extensions, some 25% (4) conflicted with their predecessors, 31% (5) offered partial support, and 44% (7) supported them.

Because of concerns about the trustworthiness of scientific findings, there currently is renewed interest in their replicability. This has occurred in fields like economics (Duvendack, Palmer-Jones, and Reed, 2017), psychology (Open Science Collaboration, 2015), statistics (Wasserstein and Lazar, 2016), and the sciences more generally (Hubbard, 2016; National Academy of Sciences, 2016). Might this lead to an increase in the publication frequency of replications? Only time will tell. Meanwhile, we must continue to estimate their incidence, the motivation for the current proposal.

Parenthetically, in judging whether a study merits replication, Makel, Plucker, and Hegarty (2012, p. 541) recommend, as they admit, the arbitrary yardstick that this should apply to articles that have been cited 100 times. According to Google Scholar, as of August 7, 2017, HA has been cited 424 times; for EBHA, this number is 214.

### (iii) Replication Plan

Since EBHA used a census of articles appearing in the *JCR*, *JM*, and *JMR* for 1990-2004, I will do likewise for the 16-year period 2005-2020. I will also add 3 more journals to those above, namely, the *Journal of the Academy of Marketing Science (JAMS), Marketing Letters (ML),* and *Marketing Science (MS),* because of their general nature and wide readership among marketing academics. So this will make my study as close as it can get to being an exact replication—faithfully repeating the methods used in the earlier work on a new sample (census in this case) from the same population (highly regarded marketing journals).

I will enlist the assistance of a Ph.D. student as a co-author in this project. In determining whether an article qualifies as a replication, we will each read and classify them all independently. This will be beneficial in three ways: (1) To help gauge the level of agreement between our two estimates of replication frequency, say using Cohen's (1960) kappa, (2) To mentor a Ph.D. student in the publication process, and (3) To inculcate in him/her a sense of the value of replication, which is sorely lacking.

To qualify as a "replication" an article has to contain an explicit citation of the original work. This does not mean that the replicating author(s) must identify their own research as such. This will be our responsibility. If doubt arises concerning whether the article is a replication, I will err on the liberal side and include it. If anything, then, the estimate of replication research will be exaggerated.

On the other hand, authors "replicating" their own work within the context of that *same* article will not count as a replication. But an author replicating his/her research in a *separate* article will be included.

Estimates of the publication frequency of replications found by EBHA will be compared with those of my follow-up at the aggregate level (e.g., *JCR* + *JM* + *JMR* versus *JCR* + *JM* + *JMR* [+ *JAMS* + *ML* + *MS*]) as well as across individual journals.

The same applies to whether the studies support, partially support, or contradict one another.

### (iv) Interpreting the Outcomes

Since the data in EBHA and my replication constitute entire populations, no statistical testing will be involved. Therefore, results will be analyzed by "eyeballing" them. For example, I would maintain that EBHA (1.2%) successfully reproduce HA's (2.4%) estimate of replication frequency because, sadly, both figures are abysmally low. Likewise, I will eyeball the numbers from this planned study with those of EBHA.

A possible explanation for the low incidence (2.4%) of replications found in HA could be a bias in favor of publishing replications contradicting (60%) earlier studies, as these may be deemed to be "original" findings. But this argument does not hold water. To begin with, as just seen, the EBHA incidence (1.2%) is smaller than HA's, as is the frequency of conflicting (25%) outcomes, while the rate of partially (31%) and fully (44%) confirmatory results is higher.

Moreover, in other disciplines the publication incidence of replications either partially or fully supporting the original results is substantial. These include forecasting at 45% and 36%; and management at 39% and 42%, respectively (Hubbard, 2016, p. 140).

Even when comparisons between studies are sample-based, which with few exceptions will be the case, I would still eyeball the outcomes to determine if a replication is successful or

otherwise. But I will not rely on *p*-values to do this because they can be extremely misleading in this context (Hubbard, 2016, pp. 70-75).

Instead, I advocate use of the criterion of overlapping confidence intervals (CIs) around parameters of interest to help decide whether a replication is successful or not because this puts the emphasis, as it should be, on comparing *effect sizes* across studies. There are, however, problems associated with the criterion of overlapping CIs. Fortunately, some of these are mitigated when accompanied with a statistical power analysis, as discussed below.

While no panacea, crucial assistance in evaluating the success or otherwise of a replication attempt can be rendered by an analysis of the statistical power exhibited in the original and replication articles. Overlapping CIs in low-powered comparisons will be unhelpful insofar as the parameter estimates are (markedly?) imprecise. As such, they afford little in the way of support for a successful replication. Indeed, one can make the case that low-powered research should not be published. And yet, as summarized by Hubbard (2016, pp. 55-56), much social science research is noticeably underpowered.

Non-overlapping CIs in low-powered comparisons can be taken as compelling evidence that the parameters from the original and follow-up studies fail to replicate. This is so despite the fact that neither of them has been estimated reliably.

Non-overlapping CIs in high-powered studies can yield mixed results concerning replication success. The parameters may or may not be from different populations. In extremely high-powered research, often purchased through huge sample sizes, CIs may not overlap because they reduce to points, even while showing strong eyeball evidence of a successful replication.

*Ceteris paribus*, overlapping CIs across two adequately- or high-powered studies constitutes meaningful *statistical* evidence of a successful replication. Adequate statistical power could be determined by adopting Cohen's (1988, p. 56) benchmark of .80.

And despite the last sentence in the above paragraph, the bottom line is this: The overlapping CI's criterion for judging the success or failure of a replication *must* be employed in a heuristic fashion, on a case-by-case basis, not as some cut-and-dried, "objective" rule. For roughly 6 decades we have become habituated to substituting the rote application of arbitrary rules-of-thumb for thinking when evaluating the worth of findings, with $p \leq .05$ as the most notorious example of this. We do not want the overlapping CIs criterion (and Bayesian equivalents) in sample-based comparisons to suffer a similar fate. So use it as another piece of information, in conjunction with others, like the theoretical and factual knowledge already accumulated in any given area.

Ah! But there's the rub. Given the publication bias against replications, there are few well-established facts/empirical regularities in marketing's literature. The upshot is a knowledge base in the discipline that is "more marsh than bedrock" (Leone and Schultz, 1980, p. 11). And yet, almost 40 years after this verdict, nothing has changed. We need ongoing emphasis on publishing replication research, and periodic assessments of its frequency and outcomes. Hence the case for the present proposal.

## References


Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement,* 20: 37-46.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences.* (2nd ed.) New York, NY: Academic Press.

Duvendack, M., R.W. Palmer-Jones, and W.R. Reed. 2017. "What Is Meant by 'Replication' and Why Does It Encounter Resistance in Economics?" *American Economic Review: Papers & Proceedings,* 107: 46-51.

Evanschitzky, H., C. Baumgarth, R. Hubbard, and J.S. Armstrong. 2007. "Replication Research's Disturbing Trend." *Journal of Business Research,* 46: 411-415.

Hubbard, R. 2016. *Corrupt Research: The Case for Reconceptualizing Empirical Management and Social Science.* Thousand Oaks: SAGE Publications, Inc.

Hubbard, R., and J.S. Armstrong. 1994. "Replications and Extensions in Marketing: Rarely Published But Quite Contrary." *International Journal of Research in Marketing,* 11: 233-248.

Leone, R.P., and R.L. Schultz. 1980. "A Study of Marketing Generalizations." *Journal of Marketing,* 44: 10-18.

Makel, M.C., J.A. Plucker, and B. Hegarty. 2012. "Replications in Psychology Research: How Often do they Really Occur?" *Perspectives on Psychological Science,* 7: 537-542.

National Academy of Sciences. 2016. *Statistical Challenges in Assessing and Fostering the Reproducibility of Scientific Results: Summary of a Workshop.* Washington, DC: National Academies Press.

Open Science Collaboration. 2015. “Estimating the Reproducibility of Psychological Science.” *Science*, 349 (6251).

Reed, W.R. 2017. “Editorial: The Practice of Replication.” *Economics E-Journal.*

Tsang, E.W.K., and K.-M. Kwan. 1999. “Replication and Theory Development in Organizational Science: A Critical Realist Perspective.” *Academy of Management Review*, 24: 759-780.

Wasserstein, R.L., and N.A. Lazar. 2016. “The ASA’s Statement on *p-V*alues: Context, Process, and Purpose.” *The American Statistician*, 70: 129-133.

Ziman, J. 1978. *Reliable Knowledge: An Exploration of the Grounds for Belief in Science.* Cambridge, UK: Cambridge University Press.

Economics
The Open-Access, Open-Assessment E-Journal

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

http://www.economics-ejournal.org/economics/discussionpapers/2017-75

The Editor

© Author(s) 2017. Licensed under the Creative Commons License - Attribution 4.0 International (CC BY 4.0).