On the conceptual foundations of replication: Reflections on R. J. Hannum, A replication plan for "Does social media reduce corruption? (2017)".

Brian D. Haig

University of Canterbury

1. Introduction

The editor of this special issue of *Economics* asked me to provide a conceptual evaluation of Randall Hannum's (2017) proposed replication plan for the primary article, "Do social media reduce corruption?' (Jha and Sarangi, 2017), along with some personal thoughts about the replication enterprise more generally. Because Hannum's thoughtful replication plan is described in very general terms, it does not lend itself to comments about the specifics of the replication process. However, his plan does contain a number of his fundamental beliefs about the nature of replication that are worth discussing.

2. The author's model

Hannum begins his article by sketching an unusual model of the replication process. The model identifies three actors: The original researcher, the replicator, and the consumer of the research findings. According to the model, it is the aim of both the original researcher and the replicator to determine what is true, and the replication is deemed to be of value if the consumer increases her certainty about the reliability of the results. I have two reservations about the usefulness of the model for casting light on the replication process. First, it seems questionable that the consumer, as an unidentified third party, could, or should, be the measure of the worth of the replication. Surely, the authors of both the original study and the replication would be the immediate judges of the worth of their findings as contributions to scientific knowledge in their sphere of competence. Beyond that, the critical scientific community (say, through the conduct of a subsequent successful meta-analysis that included both studies) would add to the earlier judgments of worth. Beyond that, consumers, such as policy advisors and decision makers, might find the results relevant to their deliberations, though it is doubtful whether they would be in a positon to judge the truth of the study outcomes.

Second, framing the model in terms of the agents' infallibility or fallibility strikes me as both unrealistic and irrelevant. It is unrealistic because infallibility is not part of the makeup of human inquirers, nor could it be. Twentieth-century studies of science by Charles Peirce, Karl Popper, and many others, make it clear that human agents are fallible. That is, their perceptual, cognitive and computational powers are limited and they, therefore, make mistakes – a point that Hannum seems to acknowledge. Consequently, the knowledge of human inquirers is provisional and subject to correction at any time. This holds for methods and strategies as well as beliefs. Error detection, then, is an important part of the scientific process, and needs to be constantly monitored. In fact, replication can be regarded as an important strategy for probing and identifying error. Because fallibilism is an inescapable feature of the human condition, I think that presenting scenarios where the replication researcher is "an infallible … discerner of the truth" and the original researcher is fallible,

and vice versa, fails to get a purchase on research reality. For this reason, I think that Hannum's model is of limited help in gauging the value of replications.

3. What is replication?

Despite widespread agreement about the importance of replication in science, there is considerable disagreement about how to understand it. Hannum refers to the fact that different typologies of replication have been proposed in an attempt to bring order to that understanding. One emerging agreement from these proposals that it is useful to distinguish, at a general level, between direct replication and conceptual replication. Direct replication duplicates the sampling and experimental procedures of the original research, whereas conceptual replication varies the salient study conditions in order to determine the scope of the original empirical findings. Most of the current debate focuses on direct replication, although it is a contested matter as to which is the more important of the two, and even whether both are needed or not. Hannum's replication plan addresses what he calls "replication with constructive scrutiny", which seeks to determine the extent to which the results are reliable. With this in mind, he reasonably asserts that a replication "can range from a quality control exercise to an expanded robustness check based on an evaluation of the data considered and the methods employed to further extensions of the original study". Given this range of possibilities, and the absence of detailed information about Hannum's replication plan, it is difficult to fathom the precise nature of the replication process, as he understands it. For what it's worth, my guess is that his plan tilts more in the direction of direct than constructive replication.

4. Tests of significance as arbiters

Despite the uncertainty in determining whether or not a replication study has been successful, most replication studies make use of tests of statistical significance to make such judgments (a point perhaps implied in Hannum's comments). As is the custom with original studies, Jha and Sarangi use tests of statistical significance to judge estimates of their model representation of the relations of social media to social corruption, but Hannum does not comment on the whether or not he thinks this is appropriate. Moreover, he does not indicate in his own proposal whether or not he would employ test of significance as an arbiter for a successful replication. In the absence of such information, I reissue a caution about using tests of significance as a measure of replication success. A number of authors have argued that it is a mistake to employ the p value as a reliable indicator of the replicability of empirical findings (Hubbard, 2016): This error has been dubbed the "replicability fantasy". The error involves the mistaken belief that the smaller the p value, the greater the evidence for replication success, as well as the belief that the complement of the p value is a direct index of replicability. A related misunderstanding is the widely held belief that a test of significance gives an acceptable measure of the probability that an effect will be replicated in a follow-up study. However, Miller (2009) has shown that these so-called replication probabilities cannot be properly estimated, and in some cases are meaningless. In any event, the use of p values as the sole criterion to judge replication success should be avoided.

5. Should all research be replicated?

In underscoring the importance of replication in science, it has been said that it is "the supreme court" of science, and that it is "the most important criterion of genuine scientific knowledge". Hannum goes further and makes the strong claim that "all research should be replicated before being accepted". However, there are a number of reasons for thinking that this requirement for the certification of scientific knowledge is too demanding. For one thing,

it could only apply to claims that are strongly empirically grounded, such as empirical generalizations. By contrast, claims about the merits of scientific theories, are more appropriately justified by invoking criteria to do with explanatory power (Haig, 2014). Even this delimitation is too stringent, for empirical claims might be justified by making use of strategies other than replication, such as calibration, triangulation, and controlling for confounds (Woodward, 1989). What matters is that a reliability, or consistency, test is used to validate a finding. Replication is the most important of these, but it is not the only one. Not all empirical research is capable of being replicated as, for example, when massive and/or longitudinal data sets are involved, making re-collecting data very difficult or impossible.

References

Jha, C. K., and Sarangi, S. (2017). Does social media reduce corruption? *Information Economics and Policy 39*, 60-71.

Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences.* Cambridge, MA: MIT Press.

Hannum, R. J. (2017). A replication plan for "Does social media reduce corruption?" (*Information Economics and Policy*, 2017). *Economics Discussion Papers,* No 2017-72, Kiel Institute for the World Economy.http://www.economics-ejournal.org/economics/discussionpapers/2017-72

Hubbard, R. (2016). *Corrupt research: The case for reconceptualising empirical management and social science.* Los Angeles, CA: Sage.

Miller, J. (2009). What is the probability of replicating a statically significant effect, *Psychonomic Bulletin & Review, 16*, 617-640.

Woodward, J. (1989). Data and phenomena. *Synthese, 79*, 393-472.