

Commentary on: Rolf A. Zwaan, Alexander Etz, Richard E., Lucas, and M. Brent Donnellan

Commentary Title: The Costs and Benefits of Replication Studies

Nicholas A. Coles

University of Tennessee

Leonid Tiokhin

Arizona State University

Anne M. Scheel

Eindhoven University of Technology

Peder M. Isager

Eindhoven University of Technology

Daniël Lakens

Eindhoven University of Technology

Abstract

The debate about whether replication studies should become mainstream is essentially driven by disagreements about their costs and benefits, and the best ways to allocate limited resources.

Determining *when* replications are worthwhile requires quantifying their expected utility. We argue that a formalized framework for such evaluations can be useful for both individual decision-making and collective discussions about replication.

In a summary of recent discussions about the role of direct replications in psychological science, Zwaan, Etz, Lucas, and Donnellan (2017; henceforth ZELD) argue that replications should be more mainstream, and discuss six common objections to direct replication studies. We believe that the debate about the importance of replication research is essentially driven by disagreements about the value of replication studies and the best way to allocate limited resources. We suggest that a decision theory framework (Wald, 1950) can provide a tool for researchers to (a) evaluate costs and benefits in order to determine *when* replication studies are worthwhile, and (b) specify their assumptions in quantifiable terms, facilitating more productive discussions in which the sources of disagreement about the value of replications can be identified.

The main goal of decision theory is to quantify the *expected utility* (the result of a cost-benefit analysis, incorporating uncertainty about the state of the world) of possible actions in order to make an optimal decision. To determine when a replication study is valuable enough to perform, we must compare the expected utility of a replication study against alternative options (e.g., performing a conceptual replication, or pursuing novel lines of research). In this commentary, we explore some of the costs and benefits of direct replications, and emphasize how different assumptions can lead to different expected-utility judgments.

Costs and benefits of direct replications

The expected utility of replication studies depends on several factors, such as judgments about the reliability of the literature, the perceived public interest in a finding, or the judged importance of a theory. Importantly, these assessments are subjective, and can lead to disagreements among researchers. Consider the concerns addressed by ZELD: Should we continue to examine highly

context-dependent effects, or limit ourselves to effects that are robust across contexts? Should we spend more resources on direct, or conceptual replications? Are direct replications prohibitively costly in large-scale observational studies? The answer is: It depends.

Highly context-dependent effects might, as ZELD note, make it “difficult if not impossible for new knowledge to build on the solid ground of previous work” (Concern I). However, to argue against pursuing these research lines, one must make the case that such costs outweigh the expected benefits. In some research areas, such as personalized medicine, highly context-dependent effects may be deemed worthwhile to pursue. If a researcher believes some (perhaps even all) effects are highly context-dependent, they should be able to argue why these effects are important enough to study, even when progress is expected to be slow and costly.

Some researchers argue that even a single replication can be prohibitively costly (Concern III). For example, Goldin-Meadow stated that “it’s just too costly or unwieldy to generate hypotheses on one sample and test them on another when, for example, we’re conducting a large field study or testing hard-to-find participants” (2016). Some studies may be deemed valuable enough to justify even quite substantial investments in a replication, which can often be incorporated into the design of a research project. For instance, because it is unlikely that anyone will build a Large Hadron Collider to replicate the studies at CERN, there are two detectors (ATLAS and CMS) so that independent teams can replicate each other’s work. Thus, high cost is not by itself a conclusive argument against replication. Instead, one must make the case that the benefits do not justify the costs.

The expected utility of a direct replication (compared to a conceptual replication) depends on the probability that a specific theory or effect is true. If you believe that many published findings are false, then directly replicating prior work may be a cost-efficient way to prevent researchers from building on unreliable findings. If you believe that psychological theories usually make accurate predictions, then conceptual extensions may lead to more efficient knowledge gains than direct replications (Concern II). An evaluation of costs might even reveal that neither direct nor conceptual replications are optimal, but that scientists should instead focus their resources on cheaper methods to increase the reliability of science (Concern IV).

The value of replication studies is also influenced by the anticipated interpretation of their outcomes (Concern VI). If we cannot reach agreement about how to evaluate a given result, its benefit to the field may be close to zero. The outcome of a replication study should increase or decrease our belief in an effect, or raise new questions about auxiliary assumptions that can be resolved in future studies. Replications may thus have higher subjective value when consensus about the interpretation of outcomes can be determined a priori (e.g. via pre-registered adversarial collaboration).

Replication attempts may also have social costs and benefits for researchers who perform replication studies, or whose work is replicated. One strength of decision theory is that it allows us to incorporate such social components in cost-benefit analyses. For example, researchers currently seem to disagree about when, and how much, reputations should suffer when findings do not replicate (Concern V). If the reputational costs of unsuccessful replications are too high, scholars may be overly reluctant to publish novel or exploratory findings. If the reputational costs are non-

existent, scholars may not exert ideal levels of rigor in their work. The social norms influencing these costs and benefits are shaped by the scientific community. Explicitly discussing those norms can help us to change them in ways that incentivize direct replications when they, ignoring the social consequences, would have high utility.

Conclusion

It is unlikely that directly replicating every study, or never directly replicating any study, is optimally efficient. A better balance would be achieved if researchers performed direct replications when the expected utility exceeded that of alternative options. Decision theory provides a useful framework to discuss the expected utility of direct replications based on a quantification of costs and benefits. A more principled approach to deciding when to perform direct replications has the potential to both help researchers optimize their behavior and facilitate a more productive discussion among researchers with different evaluations of the utility of replication studies.

References

Goldin-Meadow, S. (2016). Preregistration, replication, and nonexperimental Studies. *APS*

Observer, 29(8).

Wald, A. (1950), *Statistical decision functions*. New York: John Wiley and Sons.