

REPRODUCIBILITY IN PSYCHOLOGY:
THEORETICAL DISTINCTION OF DIFFERENT TYPES OF REPLICATIONS

Aleksandar Baucal

Department of Psychology, Faculty of Philosophy, University of Belgrade

Alex Gillespie

Department of Psychological and Behavioural Science, London School of Economics

Ksenija Krstić

Department of Psychology, Faculty of Philosophy, University of Belgrade

Tania Zittoun

University of Neuchâtel

Author Note

Correspondence author: Aleksandar Baucal, abaucal@f.bg.ac.rs

Abstract

Debates about replication in psychology have focused on methodological issues and how to strengthen the replication culture. In most cases, these discussions have tended to assume that the phenomena being investigated are universal. In this paper, we are going to propose a theoretical distinction of different types of replication. The distinction is based on the assumption that besides of universal psychological phenomena there are also phenomena, especially in social and cultural psychology, that are expected to vary between socio-cultural contexts and across history. Taking this insight to its logical conclusion it implies that the main purpose of a replication and interpretation of its results depends on the phenomenon being studied. In the case of the universal phenomena, the replication serves to validation purpose, while in the case of the socio-cultural phenomenon it serves to advance our theoretical understanding of how the given phenomenon is formatted by the socio-cultural-historical context.

Keywords: Replication in psychology; Reproducibility in psychology; Socio-cultural-historical psychology

REPRODUCIBILITY IN PSYCHOLOGY:
THEORETICAL DISTINCTION OF DIFFERENT TYPES OF REPLICATIONS

Although replicating research findings has long been discussed in psychology (Touhey 1981), there has been renewed attention since the Reproducibility Project found that only 47% of effects published in psychology journals were successfully replicated (Open Science Collaboration 2015).

Despite broad agreement that replicable findings are a cornerstone of psychology, there is a debate about the implications of replication failures. Some have suggested that alternative statistics higher replicability (Etz and Vandekerckhove 2016; Gilbert et al 2016; Hartgerink, Wicherts and van Assen 2017; Wilson and Wixted 2018). Others have proposed rules to facilitate, regulate, and enhance the transparency of replication attempts (Kahneman 2014; Nosek et al 2015).

Our contribution builds on commentaries emphasizing the contextual nature of psychological phenomena (Touhey 1981; Feldman-Barrett 2015) and research demonstrating that researchers are good at predicting replication success (Camerer et al 2018; Van Bavel et al 2016). Specifically, we propose a conceptual framework that distinguishes different types of replications and explains researchers' capability to predict replication success. The conceptual framework is based on the assumption that, in addition to universal psychological phenomena, there are also phenomena that are constituted by the socio-cultural-historical context (SCHC) in which they take place, such that we would expect them to vary between socio-cultural contexts and across

historical time (Cole 1985; Rogoff 2003; Vygotsky 1934/2012; Wertsch 1993; Valsiner 2007; Valsiner and Rosa 2015).

Although the influence of SCHCs on psychological phenomena forms a continuum, three levels of influence can be conceptualized based on interrelation between the psychological phenomena and the SCHS. First, there are '*SCHC-independent*' *psychological phenomena* that correspond to universal psychological processes (e.g. orienting reflex, basic sensory processes, perceptual learning, core attention, and memory mechanisms). Second, there are '*SCHC-dependent*' *psychological phenomena*, where the SCHC moderates the extent to which a psychological phenomenon is expressed (e.g., math competence, memory ability, attributions). Finally, at the extreme, '*SCHC-constituted*' *psychological phenomena* are qualitatively different across SCHC and might exist only in certain SCHC (e.g. specific attitudes and stereotypes, gender roles, religions, certain mnemonic strategies). It is worth noting here that meaning of 'context' is different across these three cases. In the case of SCHC-independent and SCHC-dependent phenomena the context is typically treated as categorical and fixed (e.g., country of residence or group membership) while in the case of SCHC-constituted phenomena context needs to be treated as a more dynamic, semiotic, subjective and interpretative (i.e., based on specific local meanings).

In most cases the discussion about replication has focused on SCHC-independent phenomena, but, taking account of SCHC-dependent and SCHC-constituted phenomena can help to interpret replication successes and failure and even spur theoretical innovation.

Utilizing this framework reveals three types of replication. Contrary to other replication typologies based on methodological characteristics (Hüffmeier, Mazeia and Schultze 2018), the proposed distinction is based on the hypothetical nature of psychological phenomenon itself.

Therefore, even in the case of an exact methodological replication, the interpretation of replication findings should be different for SCHC-independent and SCHC-dependent phenomena (Table 1).

INSERT TABLE 1 HERE

First, if the phenomenon being studied is assumed to be universal (SCHC-independent), then we would expect findings to be replicated regardless of the context, that is, results of an exact replication in different SCHCs should fall within the confidence interval (CI) of the original effect. If this is not the case, but results go in the same direction, then they can be used for a better meta-analytical estimation that will serve as a standard for subsequent replications. However, in the case that results vary extent across different SCHCs then it will require a reconsideration of the assumption that the given phenomenon is universal and a search for potential moderators in order to understand better the socio-cultural-historical nature of the phenomenon.

Second, if we assume that a phenomenon is SCHC-dependent in the sense that it is qualitatively invariant across SCHCs (it does not change its nature), but quantitatively it can vary depending on specific characteristics of the context, then we would not expect findings to be replicated exactly in different SCHCs, nor we would interpret a lack of exact replication as a failure. For example, gender differences in reading competence are in favor of girls in all countries participating in PISA 2012, but the gap varies across countries (OECD 2013). In fact, in such cases, we would not use replication studies to validate findings from the original study, but instead as a research tool to develop a better understanding of which aspects of the SCHC are

factors influencing the intensity of the phenomenon. This situation is quite common in cross-cultural studies (Berry et al 2011).

The third, and neglected, type of replication is related to SCHC-constituted phenomena, that is, phenomena constituted in an essential way by the specific SCHC. Such phenomena are socio-cultural-historical by their nature and, consequently, cannot be studied independently from the SCHC. In this case, even an exact replication in a different SCHC would be doubtful since it does not guarantee that the same phenomenon is studied nor that effects can be compared. Yet, if the original study and its replication are in SCHCs similar in all relevant aspects, their findings can be compared and integrated since the phenomenon is the same. However, when SCHCs in two studies are different in relevant terms then we could not treat them as a case of typical replication. In fact, regardless of using the same label for a given phenomenon and the same method, the two studies would actually be dealing with different phenomena. For example, learning school math has quite different meanings for street children and middle-class children in Brazil (Nunes, Schliemann and Carraher 1993) or implicit theories of intelligence define what it means to be clever in different SCHCs (Sternberg and Grigorenko 2004). In these cases, an exact replication in a different SCHC would not deal with the same phenomenon at all. Still, in this case, a replication provides an opportunity for the development of a theory better fitted to account for how the given phenomenon is constituted by its SCHC.

The proposed conceptual framework can explain why replication success is higher for cognitive psychology than for social psychology (Open Science Collaboration 2015) and also the finding that researchers' estimation of contextual sensitivity of phenomenon is positively related to replication successes (Camerer et al 2018; Van Bavel 2016). Specifically, social psychology rarely deals with SCHC-independent phenomena, and, consequently, we suspect that researchers

are implicitly judging the extent to which a given phenomenon might be SCHC independent when estimating replication success.

Furthermore, the proposed conceptual framework implies that researchers need to provide arguments, theoretical and/or empirical ones, on interdependency or independency of the given phenomenon on SCHC before they plan a replication. Without it, the replication can mislead rather than help to create a new knowledge, especially when dealing with a SCHC-constituted phenomenon.

The proposed framework also has practical implications. Researchers should specify the extent to which they believe the phenomena under study might vary by SCHC and provide a more substantive description of the SCHC in the method section (especially when they study socio-cultural-historical phenomena) to enable the interpretation of replication findings. In this way, the replication culture in psychology will come closer to the Replication Project claim that it might “increase certainty when findings are reproduced and promote innovation when they are not.” (p. 943) (Open Science Collaboration 2015).

References

- Berry, J. W., Poortinga, Y. H., Breugelmans, S. M., Chasiotis, A., & Sam, D. L. (2011). *Cross-Cultural Psychology: Research and Applications*. Cambridge: Cambridge University Press.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, V., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E. J., & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour* 2, 637–644, <https://dx.doi.org/10.1038/s41562-018-0399-z>.
- Cole, M. (1998). *Cultural Psychology: A Once and Future Discipline*. Cambridge: Harvard University Press.
- Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the Reproducibility Project: Psychology. *PLoS ONE*, 11: e0149794, <https://dx.doi.org/10.1371/journal.pone.0149794>.
- Feldman-Barrett, L. (2015). Psychology is not in crisis. *New York Times* (Sept. 1, 2015):A23 (2015). Available at www.nytimes.com/2015/09/01/opinion/psychology-is-not-in-crisis.html?_r=1. Accessed August 25, 2018.
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science*, 351: aad7243, <https://dx.doi.org/10.1126/science.aad7243>.

Hartgerink, C. H. J., Wicherts, J. M., & van Assen, M. A. L. M. (2017). Too good to be false:

Nonsignificant results revisited. *Collabra: Psychology*, 3(9),

<https://dx.doi.org/10.1525/collabra.71>.

Hüffmeier, J., Mazeia, J., Schultze, T. (2016). Reconceptualizing replication as a sequence of

different studies: A replication typology. *Journal of Experimental Social Psychology*, 66,

81–92, <https://dx.doi.org/10.1016/j.jesp.2015.09.009>.

Kahneman, D. (2014). A new etiquette for replication. *Social Psychology*, 45, 310-311.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S.,

Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese,

J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J.,

Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E.,

McNutt, M., Miguel, E., Levy Paluck, E., Simonsohn, U., Soderberg, C., Spellman, B.

A., Turitto, J., Van den Bos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., & Yarkoni,

T. (2015). Promoting an open research culture. *Science*, 348, 1422-1425,

<https://dx.doi.org/10.1126/science.aab2374>.

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street Mathematics and School*

Mathematics. New York: Cambridge University Press.

OECD (2013). *PISA 2012 Results: What Students Know and Can Do*. Paris: OECD.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science.

Science, 349: aac4716, <https://dx.doi.org/10.1126/science.aac4716>.

Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford: Oxford University

Press.

- Sternberg, R. J., & Grigorenko, E. L. (2004). Intelligence and culture: how culture shapes what intelligence means, and the implications for a science of well-being. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 1427–1434, <https://dx.doi.org/10.1098/rstb.2004.1514>.
- Touhey, J. C. (1981). Replication failures in personality and social psychology: negative findings or mistaken assumptions? *Personality and Social Psychology Bulletin*, 7, 593-595.
- Valsiner, J. (2007). *Culture in Minds and Societies: Foundations of Cultural Psychology*. London: SAGE.
- Valsiner, J., & Rosa, A. (2015). *The Cambridge Handbook of Sociocultural Psychology*. New York: Cambridge University Press.
- Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *PNAS*, 113, 6454-6459, <https://dx.doi.org/10.1073/pnas.1521897113>.
- Vygotsky, L. (1934/2012). *Thought and Language*. Cambridge: MIT Press.
- Wertsch, J. (1993). *Voices of the Mind: Sociocultural Approach to Mediated Action*. Cambridge: Harvard University Press.
- Wilson, B. M., & Wixted, J. T. (2018). The Prior Odds of testing a true effect in cognitive and social psychology. *Advances in Methods and Practices in Psychological Science*, 1, 186-197, <https://dx.doi.org/10.1177/2515245918767122>.

Tables

Table 1

Different types of replication distinguished based on the kind of phenomena being studied (SCHS-independent, SCHC-dependent, SCHC-constituted) and the similarity/dissimilarity of the context of original study and the context of replication study. The table provides a guide for researchers to identify different types of replication, to formulate hypothesis about replication results, and to interpret replication results.

		Replication		
		The SCHC is the same	The SCHC is partially different	The SCHC is fundamentally different
Phenomena	SCHC-independent The phenomena is both quantitatively and qualitatively invariant	Replication serves to validate a finding The finding should be reproduced exactly ¹		
	SCHC-dependent The phenomena is quantitatively variant and qualitatively invariant			
	SCHC-constituted The phenomena is both quantitatively and qualitatively variant	Replication studies identify contextual aspects and moderate findings Findings can be different ²	Replication studies develop a theory of how the SCHC constitutes the phenomenon Findings cannot be compared quantitatively	

Note: (1) The finding is reproduced exactly when it falls within the confidence interval of the original finding; (2) Findings from a replication study can be different (both in terms of direction and effect size) to the original study depending on similarities and differences between SCHCs in which they are realized.