Your Submission TEO-2021-0150R1 has been sent to production! - [EMID:623684d565d3a58c]

1 messaggio

Journal of Theoretical and Philosophical Psychology <em@editorialmanager.com> 15 giugno 2022 22:27
Rispondi a: Journal of Theoretical and Philosophical Psychology <smouton@apa.org>
A: Vera Matarese <vera.matarese@gmail.com>

Journal of Theoretical and Philosophical Psychology
TEO-2021-0150R1
Against the Resampling Account of Replication

Dear Dr. Matarese,

Your article was recently accepted to Journal of Theoretical and Philosophical Psychology and, after completion of the publication forms, has been sent along to our production team for proofing and a light copyedit. You should receive your proof within the next four weeks. Please attend carefully to the proofed copy of your work when it arrives and verify the accuracy of all data, as well as attend to any spelling, word usage, or citations that need further editing. Of course, timely dissemination of your article necessitates that you rapidly provide your corrected proof to the publisher. Following return of your proof, your article will be made available as an online first publication ahead of print.

Again, we encourage you to promote the visibility of your work when your article appears online and again when it appears in print. You can tag @APA_Journals on Twitter (or @APAJournals on Facebook) so that the publisher's social media team can easily share/retweet your posts to a broad audience - the publisher's Facebook followers mainly consist of early career practitioners, students, and international groups with a general interest in psychology.

Thank you and congratulations!

Journal of Theoretical and Philosophical Psychology

Journal of Theoretical and Philosophical Psychology

Editors: Bruce Jennings, MA, and James Cresswell, PhD

ISSN: 1068-8471

eISSN: 2151-3341

Published: quarterly, beginning in February

This journal is a publication of APA Division 24 (Society for Theoretical and Philosophical Psychology)

Journal scope statement

The *Journal of Theoretical and Philosophical Psychology*® is devoted to fostering discussion at the interface of psychology, philosophy, and metatheory. The journal addresses ontological, epistemological, ethical, and critical issues in psychological theory and inquiry as well as the implications of psychological theory and inquiry for philosophical issues.

In keeping with the journal's interdisciplinary mission, both psychology and philosophy are construed broadly to encompass a diversity of forms of inquiry such as conceptual, speculative, theoretical, empirical, clinical, historical, literary, and cultural research.

The *Journal of Theoretical and Philosophical Psychology* encourages and facilitates the informed, innovative, and critical exploration and discussion of psychological ideas and practices in both their scientific and philosophical dimensions and interrelationships.

The journal welcomes original articles, essays, and commentaries with philosophical or metatheoretical import from all disciplines concerned with human psychology.

Subscribe to the RSS feed for *Journal of Theoretical and Philosophical Psychology* (http://content.apa.org/journals/teo.rss)

Journal highlights

Announcement

New editor appointed (/pubs/journals/resources/new-journal-editors)

Call for papers

Journal of Theoretical and Philosophical Psychology

ISSN: 1068-8471
https://doi.org/10.1037/teo0000207

COMMENTARY

Against the Resampling Account of Replication

Vera Matarese
Center for Space and Habitability (CSH), University of Bern
Institute of Philosophy Institute of Philosophy, University of Bern

Edouard Machery's article "What is Replication?" deserves particular critical attention. For if it is correct, his Resampling Account of Replication has the power to reshape the current debate on replication in psychology. Indeed, with his new proposal, philosopher Machery claims to replace the "vague characterization of replication in psychology" (Machery, 2020, p. 559) with an account that deflates one of the central debates on replication—the debate which contraposes direct and conceptual replications and asks which one is preferable. In this commentary, I argue that there are deep-rooted reasons for why the distinction is meaningful, and that the Resampling Account of Replication just offers a misleading "semantic shift."

Public Significance Statement
This commentary argues against the resampling account of replication proposed in Machery (2020) and reflects on the significance of the distinction between direct and conceptual replication, which is common lore in psychology. The discussion offered in this contribution may lead to advances in the understanding of the functions and value of replication in psychology.

Keywords: direct replication, conceptual replication, Resampling Account of Replication

Machery's article "What is Replication?" deserves particular critical attention, as his Resampling Account of Replication (RAR) is meant to replace the "vague characterizations of replication in psychology" (Machery, 2020, p. 559) and psychologists' "often vague understanding of what replications are and what they are for" (Machery, 2020, p. 548). In particular, it has the ambitious goal of reshaping the current debate on replicability in psychology by deflating the usual contraposition of direct and conceptual replications.

The replication crisis, which was spurred by the discovery that many experiments in psychology are not replicable, has led, according to Machery, to "a heated controversy about the best form of replication" (Machery, 2020, p. 545).[1] In this regard, three solutions have been proposed. Some[2] suggest the superiority of direct replications:

> DirectRepl: experiments that repeat the same methodology of an original experiment, by duplicating it in all

Vera Matarese https://orcid.org/0000-0002-1644-2263
Some ideas of this article were presented during a seminar talk at Caltech University in November 2021. No data are included in this article.
Correspondence concerning this article should be addressed to Vera Matarese, Center for Space and Habitability (CSH), University of Bern, Gesellschaftsstrasse 6 (G6), Bern 3012, Switzerland. Email: vera.matarese@gmail.com

[1] This reconstruction of the debate is disputable, as it was much more articulated and nuanced than how it is presented in Machery's work. Yet, I keep the same narrative as in Machery's article, since it is the most effective way to set up this commentary. A clarification of the debate can be found in 2.4.

[2] For example, Simons (2014).

> its aspects, save for irrelevant factors such as samples, and that aim to establish its result.[3]

Others,[4] by contrast, claim that direct replications not only are too strict but even uninformative, and propose, instead, the use of conceptual replications:

> ConceptRepl: experiments that test the same hypothesis of an original experiment by changing its methodology.

A third group,[5] which I support, emphasizes the functional difference between DirectRepl and ConceptRepl and their complementary indispensability.

Philosopher Machery (2020) regards himself as the proponent of a fourth, novel position, which "reject[s] the very distinction between direct and conceptual replication as it is usually drawn" (p. 545) and claims that the debate "should be dissolved rather than resolved" (p. 545).

According to Machery (2020), "the notion of conceptual replication is confused" (p. 545). Indeed, according to him, the current debate in psychology, which distinguishes between DirectRepl and ConceptRepl on the grounds that the latter, but not the former, involves a change in methodology, rests on the naïve, even false, assumption that the methods used in an experiment are fixed factors. It is in virtue of this that, according to him, psychologists draw a distinction between those replications that change random factors (such as the samples of a population used as the experimental units) and those that change the fixed factors (the method), calling the former "direct replications" and the latter "conceptual replications."[6] However, this distinction is deceptive because, according to Machery, even the methodological components of an experiment (e.g., treatments and measurements) can be random factors. Therefore, there is no reason for differentiating those replications that resample treatments or measurements and those resampling experimental units. In light of this, Machery (2020) argues that "the controversy about […] the value of direct or conceptual replications should be dissolved" (p. 545).

In this commentary, after listing the specificities and the merits of the RAR (The RAR and Its Merits section), I address its shortcomings. First of all, contrary to Machery's analysis, I argue that the notion of ConceptRepl in psychology is not confused, as the distinction between DirectRepl and ConceptRepl should not be regarded as being based on the naïve assumption that the methodological components of an experiment cannot be random factors (Resampling Methodological Components vs. Changing Methodology section). Second, I point out that the RAR is based on a distracting "semantic shift," which makes its conclusion irrelevant to the debate on DirectRepl and ConceptRepl (A Semantic Shift section). Third, I argue that there are deep-rooted reasons, which concern problems arising in psychological scientific practice, for not endorsing Machery's semantic shift (RAR and Psychology section).

The RAR and Its Merits

Machery proposes an account of replication, the RAR, which is based on resampling. Within RAR, experiments in psychology feature different components, which are "aspects of the experiment that can be independently modified" (p. 550), and which can be either fixed factors (if their levels exhaust their relevant population) or random factors (when their levels are randomly sampled from a population).[7] Typical experimental components are experimental units, treatments, measurements, and settings. Within Machery's account, all components can be regarded as random factors, and replications are

[3] Direct and conceptual replications are defined in various ways, but my definitions summarize the descriptions commonly used and also acknowledged by Machery. Paradigmatic definitions of direct replication are as follows: "repetition of the same experimental procedure, aspiring to be comparable to an original study in all its aspect" (Hüffmeier et al., 2016, p. 82); "repetition of an experimental procedure" (Schmidt., 2009, p. 91); "attempts at reproducing a previously observed result with a procedure that provides no a priori reason to expect a different outcome […], with different samples of the same […] materials" (Nosek & Errington, 2017, p. 1). The new samples must be representative of the same population. Paradigmatic definitions of conceptual replications are "the repetition of a test of a hypothesis […] of an earlier research work with different methods" (Schmidt., 2009, p. 91), "an attempt to test the same […] hypothesis behind the original study, but which may differ substantially" (Crandall & Sherman, 2016, p. 93); "the practice of modifying an original experimental design […]" where "the resulting designs can be very different" (Romero, 2019, p. 5).

[4] For example, Stroebe and Strack (2014) and Crandall and Sherman (2018).

[5] For example, Hüffmeier et al. (2016). The authors even introduce more kinds of replication.

[6] "The usual notion of conceptual replication treats all modifications of treatment, measurements, and setting on the model of (1)," where (1) is "changing the value of a fixed factor" (Machery, 2020, p. 560).

[7] "Levels" should be understood as the states or values that a factor could assume.

those experiments that resample an original experiment's components that are treated as random factors:

> Repl(RAR): Experiment A replicates experiment B if and only if A consists of a sequence of events of the same type as B, while resampling some of its experimental components in order to assess the reliability of the original experiment.[8]

Resampling amounts to repeating the sampling procedure by creating another sample that is representative of the same population. Depending on which component is resampled, we have four different kinds of replications: experimental unit replication, treatment replication, measurement replication, and setting replication. Replications serve to check the reliability of the experiment, where "reliable" is defined as follows:

> Reliable: A token experiment is reliable if and only if, if one repeatedly sampled new values for the experimental components that are treated as random factors, everything else being kept constant, the same experimental outcome would be found with high frequency.[9]

According to Machery, given that all components of an experiment, even methodological ones, can be treated as random factors and resampled, and given that all kinds of replications play the very same role, which is to check for reliability, the distinction between direct and conceptual replications is flawed.

Let me first highlight the merits of Machery's account. His contribution is commendable for pointing out that (a) all experimental components, including methodological ones such as treatments and measurements, can be random factors and so can be represented as samples; (b) it would be unmotivated to draw any conceptual difference between replications resampling experimental units and those resampling methodological components of an experiment.

To illustrate (a), which has also been endorsed in Halina (2021), consider a simplification of a case discussed in Machery's article. Suppose that an experiment tests the hypothesis that human minds are predisposed against Bayesian inference by asking some participants to solve 15 vignettes. Not only does the experiment involve a sample of participants but also a "methodological sample," since the 15 vignettes could well be considered as a representative sample of the broad population of all vignettes featuring Bayesian problems.

To illustrate (b), it is sufficient to point out that resampling amounts to repeating the sampling procedure by creating another sample, another subset of items that is representative of its population (in our previous example, resampling would amount to choosing a different sample of the vignettes). No matter which experimental component (methodological or not) is resampled, the act of resampling serves the very same general function of checking sampling errors inherent in the sample being replaced, and, more generally, the reliability and precision of our experiment's result with regard to the experimental component that has been resampled. Therefore, it would be wrong to draw some conceptual difference between replications resampling methodological components of an experiment and replications resampling other components.

From these considerations, it follows that Machery is right in drawing two conclusions from (a) and (b), namely that (c) we ought not to presuppose which experimental components are fixed and which should be random, and that (d) we ought not to conceptually distinguish the two kinds of replication on the basis of which component is resampled. So, I will take not only (a) and (b) but also (c) and (d) as convincingly established.

Machery's points (a)–(d) are valuable especially in light of the well-known account of direct and conceptual replications in Schmidt (2009). Even though Schmidt defines the function of direct replication merely in terms of the generic function of controlling for sample errors, which would be in line with DirectRepl, he claims that this function targets the reliability of the participants' selection (experimental unit), omitting that it could serve to check the reliability of other samples, like one of methodological components. These points are also relevant for all cases where experimenters erroneously think that they do conceptual replications just because they change a methodological component without considering whether it is a fixed or random factor.

However, apart from this important contribution, several shortcomings of the article should not pass unnoticed, and, for this reason, I address them in the following sections.

[8] Machery (2020, p. 556).
[9] Machery (2020, p. 555).

Resampling Methodological Components Versus Changing Methodology

Even though (c) and (d) are correct, I argue that it is misleading to use them, as Machery does, to claim that ConceptRepl is "confused" and to conclude that the debate on the contraposition of DirectRepl and ConceptRepl should be dissolved. Indeed, apart from limited cases, the current practice of distinguishing DirectRepl from ConceptRepl is vulnerable neither to the accusation of presupposing which experimental components are fixed and which are random (violation of c), nor to that of distinguishing the two kinds of replications only on the basis of which component is resampled (violation of d).

Recall that the current debate distinguishes between replications that keep the same methodology of an experiment by keeping all aspects identical save for irrelevant factors such as samples[10] (DirectRepl), and replications changing the methodology[11] and testing the same hypothesis (ConceptRepl). Notice that according to DirectRepl any kind of resampling, regardless of which component is resampled (including treatments or measurements!) would still be counted as a direct replication, as long as the samples are representative of the same population and everything else is kept the same expect for other irrelevant factors. That is, if an experiment consists only in resampling (even treatment or measurement components) from the very same population, it is an instance of DirectRepl; if it changes methodology but checks for the very same hypothesis, it is an instance of ConceptRepl. In order to see that this distinction is meaningful, it is sufficient to demonstrate that resampling a methodological component (a form of DirectRepl) is not the same thing as changing the methodology (ConceptRepl). I will do so, by showing that (a) changing the methodology of an experiment (ConceptRepl) may not involve resampling at all, and (b) resampling a methodological component (which is a form of DirectRepl) may not involve changing the methodology of an experiment (ConceptRepl).

As for (a), consider that a change in the methodology, which can be understood as the design or the strategy that the experiment adopts in order to test a set of hypotheses (Footnote 11), may involve collecting different data, endorsing different operationalizations, employing different instruments or treatments, or utilizing different theoretical frameworks. This means that there are cases of ConceptRepl that does not involve resampling but other changes.

As for (b), consider experiments differing only with regard to their methodological samples. For instance, a replication of the experiment on Bayesian inference mentioned above consists in replacing the sample of vignettes used in the original experiment with another sample of vignettes. Recall that the resampling must be from the *same population*, and so that the two samples of vignettes must be representative of the same population of vignettes featuring Bayesian problems. Changing the sample of vignettes with another sample of vignettes (which is a form of DirectRepl) does not amount to changing the experiment's methodology (ConceptRepl): The original experiment and its replication do not use different methodologies. This means that resampling a methodological component may not involve a change in methodology.

However, this clarification may not be helpful to assess cases that involve a change in some methodological component but not in the overall methodology. In order to appreciate the appropriateness of the distinction between DirectRepl and ConceptRepl also in these cases, it is important to recognize that these two kinds of replication serve different functions. DirectRepl, by allowing only resampling from the same population, can be used to assess the precision or reliability of the experiment and informs experimenters about the presence and nature of random (in particular sample) errors, as well as sampling variability. ConceptRepl introduces novel systematic errors, which can allow for cross-checking of experimental accuracy. Since systematic errors typically involve problems with the methodological design, the type of instruments or of measurement equipment used, the most obvious way to check for systematic errors is to change the aspect of the methodology that brings that systematic error. A mere replacement of one sample with another

[10] It is assumed, like Machery does, that the new samples are representative of the same populations as the old one and so that "resampling," is from the same population.

[11] The methodology should be understood as "the experimental procedure" (Schmidt, 2009, p. 91), the "design" of the experiment (Romero, 2017, p. 4), and the specific strategy or "technique" adopted to test a set of hypotheses (Nosek & Errington, 2017, p. 1).

sample, *if they are both representative of the same population*, would generally not allow this, as a mere change of sample, *if properly randomized*, should generally not bring new systematics.[12] Suppose, instead, that a replication modifies the vignettes used in the original study by including some pictures in them or replacing them with comics. Even though this change does not affect the overall design of the experiment, it counts as a change in the methodology, thus as a ConceptRepl, because it introduces new systematic errors in the experimental design (e.g., the pictures in the vignettes or the comics could actually induce the participants to favor an intuitive answer over a reasoned one). One could (mistakenly) consider the new sample of vignettes with pictures as a "sample" of the same population as the sample of vignettes without pictures, and in light of this, one could conclude that this "resampling" constitutes an instance of DirectRepl. But this would be wrong because the change of sample does not involve a true resampling from the very same population: the new sample brings new systematics, and so cannot be regarded as equally representative of the same population as the previous one. In this case, a change of sample cannot be regarded as an irrelevant factor, and so should be rightly considered as an instance of ConceptRepl.

To conclude, the distinction drawn in the literature between DirectRepl and ConceptRepl should not be regarded, apart from some unfortunate cases, as vulnerable to Machery's criticisms, because is not based on which component of an experiment is resampled, nor on the assumption that methodological components are fixed factors and cannot be random. It is rather based on their functional difference. If the replication keeps everything the same save for samples, and so checks for precision and reliability, it is an instance of DirectRepl. When the replication involves a change in methodology, by introducing different systematic errors and thereby allowing accuracy to be checked, it counts as an instance of ConceptRepl.[13]

A Semantic Shift

Machery, however, has a reply. If "conceptual replications" do not consist in merely resampling experimental components, they should not count as replications at all. Indeed, as Repl(RAR) makes clear, replication *is* nothing more than an act of resampling experimental components that are treated as random factors. If we claim, as we did, that conceptual replications involve a change in the methodology which should not be identified as an act of resampling the methodological component of an experiment, that procedure cannot be a "replication." Rather, according to Machery (2020), it would constitute a "different experiment" (p. 557). Thus, it would not make sense to contrapose replications (Repl(RAR)) with other experiments (ConceptRepl): the debate over DirectRepl and ConceptRepl should be dissolved.

I resist this answer. Of course, ConceptRepl does not count as replication according to the RAR, where replication is defined as Repl(RAR). Note, however, that Repl(RAR) are forms of DirectRepl. This way, Machery restricts the notion of replication to only one of the two kinds of replication normally discussed in the literature, and whatever is not DirectRepl, such as ConceptRepl, will simply not count as replications. But this restriction of replication to direct replications is only a distracting *semantic shift*. If Machery does not want to label ConceptRepl as "replications," because he regards them as "different experiments," he is free to do so. But this does not change the fact that the crucial question under discussion involves precisely these experiments that Machery dismisses as "different experiments." I would push the point further. Of course, in some sense, ConceptRepl should count as different experiments from the original ones. But it is exactly in those kinds of "different experiments" that we are interested. Refusing to call them "replications" does not dissolve the debate on the preferability of ConceptRepl over DirectRepl. Machery may be skeptical: If these "different experiments" can count as conceptual replications, what would distinguish

[12] Of course, there are some cases where samples are not properly randomized or are not properly representative of their entire population, and so a change of sample may bring new systematics. See the following example.

[13] That the distinction between DirectRepl and ConceptRepl should be understood as based on a functional difference can be seen in Simons (2014), where DirectRepl is said to check the reliability of an effect (which here I refer to as "reliability" or "precision"), whereas ConceptRepl check for the truth of a theoretical hypothesis (which here I refer to as "validity" or "accuracy"). A functional distinction is also presupposed in Crandall and Sherman (2016), Nosek and Errington (2017), and Stroebe and Strack (2014).

conceptual replications from completely different experiments that have nothing to do with the original experiments we want to replicate? The answer is in the definition of conceptual replication itself: As long as these experiments test the same set of hypotheses as the original experiment, then they should count as their conceptual replications.

Machery may insist that even though the debate is normally conceived in the way just spelled out, and so between "replications" (DirectRepl) and certain "other experiments" (ConceptRepl), the only meaningful way to understand conceptual replications and the debate over them should be to conceive of conceptual replication as the resampling of methodological components treated as random factors. Indeed, the stated goal of the RAR is exactly "to provide a general account of what replication is" because "a general account of what a replication is and what it is useful for is needed to assess the respective merits of direct and conceptual replications" (Machery, 2020, p. 547). In light of this, Machery may hold that the common ground of DirectRepl and ConceptRepl can be found by repurposing conceptual replication as a kind of resampling, and that once this is done, it is clear that the debate should be dissolved because the two notions of replications have the same function.

I share the concern that motivates the RAR, and *prima facie*, by redefining any form of replication as resampling, Machery provides the common ground to establish a sort of comparison between direct and conceptual replications, by contraposing resampling of experimental unit component (direct replication) and resampling of the methodological ones such as therapies and measurements on the other hand (conceptual replications). However, on a closer look Machery's account does not achieve its goal, as it does not provide a common ground for DirectRep and ConceptRep; rather it provides a common ground for DirectRep(RAR) and ConceptRep(RAR):

> DirectRepl(RAR): act of resampling the experimental units that are treated as random factors.
>
> ConceptRepl(RAR): act of resampling an experiment's methodological experimental components (such as treatments or measurements) that are treated as random factors.

His account provides a common ground for direct and conceptual replications only insofar as we accept his semantic shift, regard replication only as Repl(RAR), reject ConceptRepl as a kind of "replication," and embrace a new definition of conceptual replication as ConceptRepl(RAR). Naturally by doing this, the debate on the superiority between direct and conceptual replications reduces to an insignificant contraposition between ConceptRepl(RAR) and DirectRepl(RAR), both of which are simply instances of DirectRepl.

In sum, while Machery is perfectly within his rights to endorse the semantic shift, his account merely shows that given his own definitions the debate on direct and conceptual replications should be dissolved because either it should be understood as contraposing replications against other kinds of experiments or as a contraposition of different kinds of resampling which have the very same function. It by no means shows that the debate in the literature should be dissolved.

RAR and Psychology

Machery could reply that at the core of his account is not a distracting semantic shift but a piece of constructive conceptual engineering, based on what psychologists actually do (Machery, 2020, p. 548) and on the problems they deal with. In particular, he may support this point by arguing that (a) the contraposition of DirectRepl and ConceptRepl does not reflect scientific practice, and that (b) replications within psychology have the role to check for reliability only.

I start with claim i. This would be right in case the controversy between direct and conceptual replications were the result of idle intellectual speculation, or if it were not grounded in cases involving ConceptRepl. By contrast, I argue that the debate is grounded in problems concerning the use of ConceptRepl and which are inherent in psychological practice. Moreover, since these problems would remain untouched by Machery's "conceptual engineering," we have good reasons for not endorsing his semantic shift.

The preferability of direct or conceptual replications is particularly debated within social psychology, whose theories rely on iconic studies that are hardly directly replicable because of cultural and social changes. Stroebe and Strack (2014) describe an experiment conducted in 1959, where female students were asked to read aloud "12 obscene words" and "two vivid descriptions of sexual activity" (p. 61). The experiment would certainly have a different result if repeated with today's female students, as it

would probably trigger "amusement rather than embarrassment" (p. 61). Therefore, what is needed is conceptual replication, that is, the experiment needs to be redesigned. Stroebe and Strack further discuss the case of the dual-process theories of persuasion, according to which the impact of the quality of the arguments is greater the more thoughtfully and deeply the communication is processed by the recipients. This thesis was confirmed by employing different experimental designs, such as distracting the recipients, giving the impression that the real discussion would happen in a future meeting, or else by pretending that there was a need for immediate closure.

The controversy on the best form of replication arises from this and similar cases of scientific practice, which point to the practical limitations. In the case concerning the unfeasibility of DirectRepl, the problem was whether the theories of social psychology can rely heavily on conceptual replications without losing credibility. To be noticed is that the case involves ConceptRepl rather than ConceptRepl(RAR), as the replications involve several changes regarding the overall experimental design. Each new method introduces different systematic errors and so cannot be reduced to a mere resampling procedure. Therefore, the debate on the best form of replication should consider ConceptRepl, as it is correctly done in the literature, and not ConceptRepl(RAR). This shows that while the debate on DirectRepl versus ConceptRepl is grounded in psychological practice, Machery's piece of "conceptual engineering," which rejects the distinction between DirectRepl and ConceptRepl, fails to come to grips with the concrete problems that this debate is meant to solve.

One may argue that this kind of problems is accounted for within RAR not in terms of contrasting DirectRepl(RAR) and ConceptRepl (RAR) but in terms of the contraposition between Repl(RAR) and extensions. Indeed, within the RAR, the alternative to replications are extensions, which are those experiments that either change the level of a fixed factor or resample a random factor from a different population. Therefore, given that one cannot put ConceptRepl under the umbrella of Replication(RAR), within RAR, one seems to be forced to identify ConceptRepl with extension. This would be perfectly fine if extensions were experiments testing the very same hypothesis as the original one. The problem is that extensions are usually meant to test a more general hypothesis and so reach a more general conclusion, or at least to check to what extent a certain result remains invariant within a range. As Machery stresses, their role is to test validity *as well* as the invariance range of a phenomenon (p. 563). By contrast, ConceptRepl is usually employed to check the validity of a result without the intention of checking for its invariance range. Indeed, they test the very same hypothesis as a previous study and so its conclusion is the same conclusion as the one of the original experiment. One can see this by going back to the case of the ConceptRepl that involves the two samples of vignettes, one with pictures and one without. Once we know that both samples of vignettes are methods that can check for the very same hypothesis, which is that human minds are predisposed against Bayesian inferences, changing one sample of vignettes with another has the aim to cross-check the results by bringing different systematics and establish their validity. The aim of the study is not to check to what extent the result can be generalized, as we know already the two methods are interchangeable, given that they were built to provide the same result. Of course, one could always *interpret* cases of ConceptRepl as cases of extensions, but the difference between extension and ConceptRepl becomes clearer once one specifies the hypothesis that they are meant to test. ConceptRepl is meant to test the same hypothesis as a previous experiment, and not a more general one (or a different one from which one may generalize).

In this sense, the RAR is impoverished because the original concept of conceptual replication, which is to test the very same hypothesis by checking the validity of a previous result by changing the methodology and so by adding new systematics does not find its proper room. Either it becomes merely an act of resampling the methodological component treated as a random factor, but then it checks only for the reliability of the result, or it becomes an extension, which does check for validity, but whose aim is also to check to what extent the result can be *generalized* or at least *extended*.

Finally, before ending this commentary, it seems that this is the appropriate place to stress the functional pluralism of replication, by addressing claim ii. Narrowing down the functions of replication to the reliability, which, in the RAR account, merely consists in checking

for sampling error, does not stand the scrutiny of scientific psychological practice, which rather relies on the diverse functions of replication. Taking a broad view, one of the most important questions in psychology is to what extent the investigative findings of a study are reliable and valid and can be used to support a certain line of theorizing. In order to check this, replications play a central role thanks to their diverse functions. In this regard, the functional account by Schmidt is commendable for providing a complete and fine-grained typology of replication that identifies a general function of replication, which is to verify a piece of knowledge, and several more specific functions: (a) To control for sampling error (chance result), (b) to control for artifacts (lack of internal validity), (c) to control for fraud, and (d) to verify the underlying hypothesis of the earlier experiment. This approach is particularly helpful because it allows distinguishing between direct replications and conceptual replications through their functions. Indeed, while DirectRepl serves Functions a, b, c, and ConceptRepl serves Function d. Schmidt's classification is also specific enough to distinguish between ConceptRepl, which test the very same hypothesis, and what is generally referred to as extensions, which aim to extend or generalize the result.

Conclusion

Although I agree with Machery that we should neither assume which experimental components are fixed and which are random factors, nor distinguish kinds of replicability on the basis of which component is resampled, and although I agree that a general account of replication is needed, I find the RAR account wanting. First of all, the difference between DirectRepl and ConceptRepl is not based, as Machery argues, on the assumptions of which experimental components are random factors and which cannot be. Second, and most importantly, the RAR is based on a misleading semantic shift, rather than on an attentive examination of scientific practice, in particular of the specific social psychology experiments, which ground the debate that the RAR aims to dissolve.

References

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, *66*, 93–99. https://doi.org/10.1016/j.jesp.2015.10.002

Halina, M. (2021). Replications in comparative psychology. *Animal Behavior and Cognition*, *8*(2), 263–272. https://doi.org/10.26451/abc.08.02.13.2021

Hüffmeier, J., Mazei, J., & Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology*, *66*, 81–92. https://doi.org/10.1016/j.jesp.2015.09.009

Machery, E. (2020). What is a replication? *Philosophy of Science*, *87*(4), 545–567. https://doi.org/10.1086/709701

Nosek, B. A., & Errington, T. M. (2017). Making sense of replications. *eLife*, *6*, Article e23383. https://doi.org/10.7554/eLife.23383

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100. https://doi.org/10.1037/a0015108

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*(1), 59–71. https://doi.org/10.1177/1745691613514450

Received October 31, 2021
Revision received June 1, 2022
Accepted June 7, 2022 ■