Replication, Reproduction and Re-analysis: Three ways for verifying experimental findings

Omar S. Gómez G.
Universidad Politécnica de Madrid
Boadilla del Monte 28660
Madrid, Spain
ogomez@ieee.org

Natalia Juristo
Universidad Politécnica de Madrid
Boadilla del Monte 28660
Madrid, Spain
natalia@fi.upm.es

Sira Vegas
Universidad Politécnica de Madrid
Boadilla del Monte 28660
Madrid, Spain
svegas@fi.upm.es

ABSTRACT

The verification of a previously observed finding is important. Being able to verify a finding provides some guarantee that it is not product of chance but that the observed phenomenon is stable or regular. The most well done way of verifying experimental findings is through replication. But, is this the only method for verifying the observations of an experiment? We have studied different ways of verifying experimental findings that are used in other disciplines. From the 18 replication classifications that we located, we have identified three methods for verifying a finding. Each of these methods fulfil a particular verification purpose.

Categories and Subject Descriptors

D.2.0 [**Software Engineering**]: General

General Terms

Experimentation

Keywords

Experimental findings, Verification, Re-analysis, Replication, Reproduction, Experimental Software Engineering

1. INTRODUCTION

The verification of experimental findings plays a key role in scientific progress, as it consolidates knowledge. Philosopher Karl Popper [28] emphasizes its value and declares: "We do not take even our own observations quite seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated coincidence, but with events which, on account of their regularity and reproducibility, are in principle intersubjectively testable".

In Experimental Software Engineering (ESE) we have tried to verify findings through the repetition or replication of previous experiments. In the trials to synthesize the findings of a set of replicated experiments [10, 22, 23, 9], researchers report that replication results are so variable that they are hard to synthesize. This has led researchers to consider other ways of verification. As Miller [24] claims "perhaps the field should start considering that [other] alternatives may prove to be more fruitful in many replication situations".

We have looked at how other disciplines verify experimental findings with the aim of analyzing whether we can adapt these methods to our field.

The remainder of the article is structured as follows. In Section 2 we describe the three methods of verification that we have found. In Section 3 we try to clarify the concepts of *replication* and *reproduction*. In Section 4 we describe *re-analysis*, *replication* and *reproduction* as methods of verifying experimental findings in ESE. Annex A lists the descriptions of the replication types that we have found.

2. VERIFYING EXPERIMENTAL FINDINGS IN OTHER DISCIPLINES

In order to identify classifications of replications in other disciplines we searched in Google®, Google scholar®, ScienceDirect® and JSTOR® using different keywords (*types of replications*, *types of experimental replications*, *typology of replications*, *replication types*, *replication typologies*, *replication types* and *classification of replications*). After running the searches on the four engines and examining the results returned, we located an initial set of 10 classifications [33, 7, 15, 3, 32, 25, 5, 20, 12, 31]. Following the references in this initial set, we were able to locate another 8 [21, 11, 8, 18, 16, 19, 26, 29]. We ended up with 18 replication classifications shown in Annex A.

Examining the classifications, we found that findings were not always verified by running the experiment over again. Neither did the replication always repeat the experimental procedure of the baseline experiment. We have identified three major groups of methods for verifying findings:

1. Follow the **same method** used in the baseline experiment. The degree of similarity between the new experiment and the baseline experiment vary. For verification purpose some of the structural components of the experiment can be changed or modified in the replication. For example, Tsang and Kwan use the term *empirical generalization* when the study is repeated on

different populations. Monroe uses the term *independent replication* when the study is repeated by different researchers.

This type of replication is used for different purposes. According to Lykken, for example, the purpose of *operational replication* is to check that the experimental recipe produces the same results with another researcher. Tsang & Kwan's *empirical generalization* purpose is to test the extent to which the study results are generalizable to other populations.

Most researchers use the term *replication* accompanied by an adjective to refer to this method of verification, e.g. *real replication*, *strict replication*, *close replication*. The adjective denotes the change made to the structure of the experiment. Table I shows the replication types in this category.

Table I: Using the same experimental method.

Term	Author(s)
Close Replication	Lindsay & Ehrenberg
Conceptual Replication	Hunter; Monroe
Demonstrated Replication	Monroe
Differentiated Replication	Lindsay & Ehrenberg
Direct Replication	Schmidt; Kantowitz et al.
Empirical Generalization	Tsang & Kwan
Exact Replication	Ijzendoorn; Tsang & Kwan;
Experimental Replication	Leone & Schultz
Generalization and Extension	Tsang & Kwan
Independent Replication	La Sorte; Monroe
Instrumental Replication	Kelly et al.
Literal Replication	Lykken; Kelly et al.
Nonexperimental Replication	Leone & Schultz
Nonindependent Replication	Monroe
Operational Replication	Lykken; Kelly et al.
Partial Replication	Hendrick; Monroe
Real Replications	Evanschitzky & Armstrong
Reproducibility of an experiment under a fixed theoretical interpretation	Radder
Reproducibility of the material realization of an experiment	Radder
Retest Replication	La Sorte
Scientific Replication	Hunter
Sequential Replication	Monroe
Statistical Replication	Hunter
Strict Replication	Hendrick; Monroe
Systematic Replication	Kantowitz et al.; Finifter
Types 0, I, II	Easley et al.
Types A..H	Bahr et al.
Varied Replication	Ijzendoorn
Virtual Replication	Finifter

2. Use a **different method** to the baseline experiment. In this type of verification, the only thing the new experiment has in common with the baseline experiment is that they are both based on the same theoretical structure, i.e. they share the same constructs. This verification is used to corroborate previously observed findings through a different path. Hendrick, Schmidt and Kantowitz et al. call this type of verification *conceptual replication*, whereas Finifter names it *systematic replication*. Radder, describes it as the *reproducibility of the result of an experiment*. Table II shows the replication types that adhere to this verification method.

3. Use **existing data sets** from a previous experiment to reanalyse the data employing either the same analysis procedures or others. This modus operandi is useful for verifying whether errors were made during the data analysis stage or whether the outcomes are affected by any particular data analysis technique. Some replication types reanalyse the statistical models instead of the existing study data. Different names are used for this type of verification. For example, La Sorte calls it *internal replication*; Finifter terms it *pseudoreplication*, and Tsang and Kwan describe it as *checking of analysis* and *reanalysis of data*. Table III shows the replication types we identified that fall into this category.

Table II: Using a different experimental method.

Term	Author(s)
Conceptual Extension	Tsang & Kwan
Conceptual Replication	Hendrick; Schmidt; Kantowitz et al.
Constructive Replication	Lykken; Kelly et al.
Corroboration	Leone & Schultz
Differentiated Replication	Lindsay & Ehrenberg
Generalization and Extension	Tsang & Kwan
Reproducibility of the result of an experiment	Radder
Sistematic Replication	Finifter
Theoretical Replication	La Sorte
Type III	Easley et al.
Types I..P	Bahr et al.

Table III: Reanalyzing existing data.

Term	Author(s)
Checking of Analysis	Tsang & Kwan
Complete Secondary Analysis	Ijzendoorn
Data Re-analyses	Evanschitzky & Armstrong
Internal Replication	La Sorte
Pseudoreplication	Finifter
Reanalysis of Data	Tsang & Kwan
Restricted Secondary Analysis	Ijzendoorn
Types I, II	Mittelstaedt & Zorn

There are no standardized terms for identifying the three types of verification observed in the replication classifications. Searching names to identify each of the three types of verification is easy for the third one but more complex for the first two. Refer to the third as *re-analysis* makes sense since the descriptions clearly allude to this concept. However, the naming of the other two types cause some confusion.

3. REPLICATION VS. REPRODUCTION

Most researchers use the term *replication* to refer to the repetition of an experiment, although some use the term *reproduction* or *reproducibility* to describe this repetition. This would mean that many researchers consider the two terms to be synonyms. Likewise, Wikipedia uses these terms indistinctly and defines reproducibility as "one of the main principles of the scientific method, and refers to the ability of a test or experiment to be accurately reproduced, or replicated, by someone else working independently"[34].

Some researchers, however, do make a distinction between the two terms. Cartwright [4], for example, suggests *replicability* "doing the same experiment again" from *reproducibility* "doing a new experiment". For Cartwright the replication of an experiment refers to repeating an experiment very closely following the method used in the baseline experiment, whereas reproduction refers re-examining a previously observed finding using a different method to what was employed in the previous experiment.

According to Cartwright, replication does not guarantee that the observed finding corresponds to an existing regularity. The finding can be artifactual, i.e. a product of the materials or the instruments used in the experiment. To guarantee that the finding corresponds to a regularity in the reality under observation, a reproduction using different methods needs to be undertaken to ensure that the finding is independent of the procedure, materials or instruments used in the experiment.

When the results are repeatable using the same experimental method, the experimenters can be confident that they have observed some sort of regularity that is stable enough to be observed more than once. But, as it was observed using the same method, there could be a cause-effect relationship between the method and the observation. As Radder puts it [30], "[this result] does not imply any agreement about what the phenomenon is. Some interpreters may even argue that the phenomenon is an artifact, because, though it is stable, it is not to be attributed to the object under study but to certain features of the apparatus[1]". Cartwright [4] claims that "reproducibility, then, is a guard against errors in our instruments" in such a situation. According to Cartwright, though, reproduction is not absolutely necessary, as the better designed the instruments (apparatus) are, the less likely it is to have to use reproducibility.

Reproduction can be seen as a sort of triangulation, where the experimenters use different methods or procedures in an attempt to validate or corroborate the findings of the previous experiment [6]. According to Park [27], "These triangulation strategies can be used to support a conceptual finding, but they are not replications of any degree".

The concept of *replication* given by Cartwright would fit the first type of verification described in section 2, whereas the concept of *reproduction* adheres to the second type of verification.

To conclude this section, we find that in fields related to measurement, the term *reproduction* is also used. In Biochemistry, for example, a *coefficient of reproducibility* is commonly used to evaluate the accuracy of the measurements taken across more than one trial (sample). In other words, *reproduction* indicates the reproducibility of a measurement. According to Hunter [13], reproducibility refers to measures of inter-laboratory variability. This is equivalent to the definition in ISO 5725-1 [1], where, in reference to measurements, *reproducibility conditions* are defined as conditions where test results are obtained with the same method on identical test material in different laboratories with different operators using different equipment. There is a longitudinal multiple-case study that uses this coefficient in Empirical SE [2].

In Computational Science there is an initiative called *reproducible research* [17]. Reproducible research aims that anything in a scientific paper should be reproducible by the reader, including results, plots and graphs. The ultimate product should not be a published paper but rather the entire environment used to produce the results in the paper (data, software, etc.).

4. VERIFYING FINDINGS IN ESE

We propose to use the terms *re-analysis*, *replication* and *reproduction* for identifying the three types of verification. All the three types are necessary to guarantee that an experimental finding identifies regular events independent of the method used to observed them.

In **re-analysis**, the data of a previously run experiment are used to verify the results rather than re-running the experiment. Re-analysis verifies two issues:

1. Data are analyzed by other researches using the same data analysis technique to verify that no errors were made during the data analysis phase.

2. Data are analyzed with other analysis techniques to verify whether similar findings can be obtained using the same data of a previous experiment.

For example, one of the metrics in the experiment run by Kamsties and Lott [14] to measure the effectiveness of several defect detection techniques is the percentage of total possible failures observed. If we had access to the data of this experiment, we could verify the result using the same analysis procedures used by the experimenters (ANOVA) or we could re-analyse the results using a Generalized Linear Model with a logit link function.

Replication verifies that the observed findings are stable enough to be discovered more than once. Replication uses the same method as in the baseline experiment. Different types of replications need to be perform to verify the findings independence of every component of the experimental structure.

Continuing with the above example, an experimenter can vary the type of experimental subjects (using practitioners instead of students) or the artefacts (programs, source code or specifications). Each change leads to a replication type. Each replication type fulfils a purpose in the verification of the findings. The changes made in the example aim to verify whether the observed event is a local phenomenon for a particular subject type or for particular artefacts or whether it exhibits a more global pattern.

Reproduction verifies that the findings are not to be attributed to the experimental method. In reproduction a new experiment is run (using different experimental methods) to test the same hypotheses as the baseline experiment. This form of verification is to be used it is suspected that the findings observed in an experiment are artifactual (product of the apparatus). In this type of repetition, the replicator has "nothing more than a clear statement of the empirical fact" [21] which the previous experimenter claims to have established.

Again using the above example, supposing that the experimenters have observed that Technique A is more effective than Technique B with inexperienced (junior) subjects. A reproduction would mean the experimenters preparing their own materials and artefacts, and defining their own metrics to measure the response variables (i.e. creating their own apparatus). The constructs are operationalized differently to verify that the findings are independent of the experimental method used.

5. ACKNOWLEDGMENTS

This work has been performed under research grant TIN-2008-00555 of the Spanish Ministry of Science and Innovation, and research grant 206747 of the México's National Council of Science and Technology (CONACyT).

6. REFERENCES

[1]The term apparatus refers to the instruments, materials or methods used.

[1] ISO 5725-1:1994. Accuracy (trueness and precision) of measurement methods and results, ISO, Geneva, Switzerland, 1994.

[2] B. C. Anda, D. I. Sjøberg, and A. Mockus. Variability and reproducibility in software engineering: A study of four companies that developed the same system. 35(3):407–429, 2009.

[3] H. M. Bahr, T. Caplow, and B. A. Chadwick. Middletown III: Problems of replication, longitudinal measurement, and triangulation. *Annu. Rev. Sociol.*, 9(1):243–264, Aug. 1983.

[4] N. Cartwright. Replicability, reproducibility, and robustness: Comments on harry collins. *History of Political Economy*, 23(1):143–155, Mar. 1991.

[5] R. Easley, C. Madden, and M. Dunn. Conducting marketing science: The role of replication in the research process. *Journal of Business Research*, 48(1):83–92, April 2000.

[6] S. Easterbrook, J. Singer, M. Storey, and D. Damian. Selecting empirical methods for software engineering research. In F. Shull, J. Singer, and D. I. Sjøberg, editors, *Guide to Advanced Empirical Software Engineering*. Springer, London, 2008.

[7] H. Evanschitzky and J. S. Armstrong. Replications of forecasting research. *International Journal of Forecasting*, In Press, Corrected Proof, 2009.

[8] B. Finifter. The generation of confidence: Evaluating research findings by random subsample replication. *Sociological Methodology*, 4:112–175, 1972.

[9] J. Hannay, T. Dybå, E. Arisholm, and D. Sjøberg. The effectiveness of pair programming: A meta-analysis. *Information and Software Technology*, 51(7):1110–1122, July 2009.

[10] W. Hayes. Research synthesis in software engineering: A case for meta-analysis. In *Proc. of the 6th International Symposium on Software Metrics (METRICS '99)*, page 143, Washington, DC, USA, 1999.

[11] C. Hendrick. Replications, strict replications, and conceptual replications: Are they important? In J. W. Neuliep, editor, *Handbook of Replication Studies in the Social and Behavioral Sciences*. Sage, Newbury Park, California, 1990.

[12] J. Hunter. The desperate need for replications. *Journal of Consumer Research*, 28(1):149–158, 2001.

[13] J. S. Hunter. The national system of scientific measurement. *Science*, 210(4472):869–874, 1980.

[14] E. Kamsties and C. M. Lott. An empirical evaluation of three defect-detection techniques. In *Proc. of the 5th European Software Engineering Conference*, pages 362–383, London, UK, 1995.

[15] B. H. Kantowitz, H. L. Roediger III, and D. G. Elmes. *Experimental Psychology*. Wadsworth Publishing, second ed. edition, 1984.

[16] C. Kelly, L. Chase, and R. Tucker. Replication in experimental communication research: an analysis. *Human Communication Research*, 5(4):338–342, 1979.

[17] J. Kovacevic. How to encourage and publish reproducible research. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007 (ICASSP 2007)*, volume 4, pages IV–1273–IV–1276, 2007.

[18] M. A. La Sorte. Replication as a verification technique in survey research: A paradigm. *The Sociological Quarterly*, 13(2):218–227, 1972.

[19] R. Leone and R. Schultz. A study of marketing generalizations. *The Journal of Marketing*, 44(1):10–18, 1980.

[20] R. M. Lindsay and A. S. C. Ehrenberg. The design of replicated studies. *The American Statistician*, 47(3):217–228, 1993.

[21] D. T. Lykken. Statistical significance in psychological research. *Psychol Bull*, 70(3):151–159, 1968.

[22] J. Miller. Can results from software engineering experiments be safely combined? In *Proc. of the 6th International Symposium on Software Metrics (METRICS '99)*, page 152, Washington, DC, USA, 1999.

[23] J. Miller. Applying meta-analytical procedures to software engineering experiments. *J. Syst. Softw.*, 54(1):29–39, 2000.

[24] J. Miller. Replicating software engineering experiments: a poisoned chalice or the holy grail. *Information and Software Technology*, 47(4):233–244, Mar. 2005.

[25] R. Mittelstaedt and T. Zorn. Econometric replication: Lessons from the experimental sciences. *Quarterly Journal of Business & Economics*, 23(1), 1984.

[26] K. B. Monroe. Front matter. *The Journal of Consumer Research*, 19(1):i–iv, 1992.

[27] C. L. Park. What is the value of replicating other studies? *Research Evaluation*, 13:189–195(7), December 2004.

[28] K. Popper. *The Logic of Scientific Discovery*. Hutchinson & Co, 1959.

[29] H. Radder. Experimental reproducibility and the experimenters' regress. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1:63–73, 1992.

[30] H. Radder. *In and about the world : philosophical studies of science and technology / Hans Radder*. State University of New York Press, Albany, N.Y. :, 1996.

[31] S. Schmidt. 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, 2009.

[32] E. Tsang and K.-M. Kwan. Replication and theory development in organizational science: A critical realist perspective. *The Academy of Management Review*, 24(4):759–780, 1999.

[33] M. H. van IJzendoorn. A process model of replication studies: on the relation between different types of replication. *Leiden University Library*, 1994.

[34] Wikipedia. Reproducibility — wikipedia, the free encyclopedia, 2009. [Online; accessed 9-September-2009].

ANNEXES

A.

This section lists the descriptions of the located replication classifications.

Table IV: Descriptions of replication classifications.

Bahr et al.	Types A..P	This classification categorizes replications according to four dichotomic properties (equal or different) of a replication. These properties are: time, place, subjects and methods. Based on combinations of these properties, Bahr et al. define 16 replication types.
Easley et al.	Type 0 (Precise Duplication)	This replication is defined as a precise duplication of a prior study. Therefore, Type 0 (precise duplication) studies are those studies in which every nuance of the experimental setting is precisely reproduced; as such, the cause-effect relationship is finite. The ability to conduct a Type 0 replication is limited to experimenters in only some of the natural sciences. As others have stated, it is an impossibility to conduct a Type 0 replication in a social science context because uncontrolled extraneous factors have the potential to interact with the various components in an experimental setting. For example, human subjects cannot be precisely duplicated. A social scientist is limited only to matching subjects as closely as possible.
	Type I (Duplication)	A type I replication is a faithful duplication of a prior study and, as such, is considered the “purest” form of replication research in the social sciences. It should be mentioned at this point that a Type I replication is the one most closely associated with the term “replication” in the minds of most researches. More over, this is also the type of replication research most criticized for not being creative. This is somewhat ironic, given the apparent receptivity of reviewers to cross-cultural research that, in many cases, is usually the study of the generalizability of findings from a single country or culture to others and, thus, is simply a Type I replication.
	Type II (Similar)	A type II replication is a close replication is a close replication of a prior study, and a Type III replication is a deliberate modification of a prior study. Type II replications are the most common form of replication research in marketing settings and are useful in testing phenomena in multiple contexts. If effects are shown in a variety of testing contexts, the case for the findings is strengthened. This has been called the process of triangulation.
	Type III (Modification)	This replication is a deliberate modification of a prior study. In a Type III replication, the threat of extraneous factors inherent to the nature of human subjects, unless explicitly accounted for in theory testing, is not a factor of concern with regard to replicability.
Evanschitzky & Armstrong	Real Replications	This replication is a duplication of a previously published empirical study that is concerned with assessing whether similar findings can be obtained upon repeating the study. This definition covers what are variously referred to as "exact," "straight," or "direct" replications. Such works duplicate as closely as possible the research design used in the original study by employing the same variable definitions, settings, measurement instruments, analytical techniques, and so on.
	Model Comparisons	This replication is an application of a previously published statistical analysis that is concerned with assessing whether a superior goodness-of-fit can be obtained, comparing the original statistical model with at least one other statistical model.
	Data Re-analyses	This replication can be defined as an application of previously published data that is concerned with assessing whether similar findings can be obtained using a different methodology with the same data or a sub-sample of the data.
Finifter	Virtual Replication	The intention is to repeat an original study not identically but “for all practical purposes” to see whether its results hold up against chance and artifact. Virtual replications are also frequently conducted to find out how dependent a result is on the specific research conditions and procedures used in an original study. To answer this question, one or more of the initial methodological conditions is intentionally altered. For example, a survey or experiment might be repeated except for a change in measuring devices, in the samples used, or in research personnel. If the initial result reappears despite changes, faith in the original finding mounts.
	Systematic Replication	The emphasis in systematic replication is not on reproducing either the methods or the substance of a previous study. Instead, the objective is to produce new findings (using whatever methods) which are expected by logical implication to follow from the original study being replicated. When such an implication is actually confirmed by systematic replication, confidence is enhanced not only in the initial finding that prompted the replication but also both in the derived finding and in whatever theoretical superstructure was used to generate the confirmed inference.
	Pseudoreplication	It can be defined according to three main operational variations: the repetition of a study on certain subsets of an available total body of real data; the repetition of areal data study on artificial data sets which are intended to simulate the real data; and the repeated generation of completely artificial data sets according to an experimental prescription.
Hendrick	Strict Replication	An exact, or strict, replication is one in which independent variables (treatments) are duplicated as exactly as possible. That is, the physical procedures are reinstituted as closely as possible. It is implicitly assumed that contextual variables are either the same as in the original experiment, or are irrelevant.
	Partial Replication	A partial replication is some change (deletion or addition) in part of the procedural variables, while other parts are duplicated as in the original experiment. Usually some aspect of the procedures is considered “unessential”, or some small addition is made to expedite data collection.
	Conceptual Replication	A conceptual replication is an attempt to convey the same crucial structure of information in the independent variables to subjects, but by a radical transformation of the procedural variables. In addition, specific task variables are often necessarily changed as well.
Hunter	Statistical Replication	For statistical replications as perfectly replicated studies: a) All studies measure the independent variable in exactly the same way. b) All studies measure the dependent variable in exactly the same way. c) All studies use exactly the same procedure. d) All studies draw samples from the same population.
	Scientific Replication	For scientific replications for simple causal studies: a) All studies measure the same independent variable X. b) All studies measure the same dependent variable Y. c) All studies use essentially the same procedure. d) All studies should sample from populations that are equivalent in terms of the study question and hence the study outcome. The difference is that statistical replications assume that the word “same” means identical, while scientists interpret the word “same” to mean equivalent.
	Conceptual Replication	This replication verifies one of the hypotheses that were not tested in the original study. The researcher of the original study defines control groups to test the most obvious alternative hypotheses against administrative details that are thought to be irrelevant. Any treatment, intervention or manipulation is a set of administrative procedures, which are mostly intrinsic to the active ingredient of the treatment. These replications examine whether the administrative procedures influence the treatments as reflected in the dependent variable.

Table V: Descriptions of replication classifications (continued).

Ijzendoorn	Complete Secondary Analysis	It is a kind of replication in which all parameters except the researcher and the method of data analysis are kept constant. Secondary analysis also is one of the most inexpensive and efficient types of replication, because it is based on existing data sets. One of the main barriers to secondary replication is, however, the accessibility of the original data sets. The complete secondary analysis may include recoding of the original raw data. In this replication, there are two phases of processing the raw data involved: the coding and analyzing of the data.
	Restricted Secondary Analysis	In this type, the coding system is not changed but only the methods of analyzing the data, to see whether the original results survive statistical criticism or the application of refined methods of statistical analysis.
	Exact Replication	A replication will be called "exact" if it is essentially similar to the original study. This replication is applied to (dis-)confirm the doubts, and to check the assumptions of the varied replications. Many scientists feel that exact replications may be carried out, but usually are irrelevant for scientific progress.
	Varied Replication	Replications should be carried out in which new data under different conditions are being collected. From the start, the original study will be "trusted" so much that rather significant variations in the design will be applied. Larger variations may lead to more interesting discoveries in addition to the original study, but they will be followed by smaller variations if more global replications fail to produce new "facts". If even modest variations fail to reproduce the results, a more or less exact replication is needed.
Kantowitz et al.	Direct replication	This is the attempt to repeat the experiment as closely as is practical, with as few changes as possible in the original method.
	Systematic replication	The experimenter attempts to vary factors believed to be irrelevant to the experimental outcome. If the phenomenon is not illusory, it will survive these changes. If the effect disappears, then the researcher has discovered important boundary conditions on the phenomenon being studied.
	Conceptual replication	One attempts to replicate a phenomenon, but in a way radically different from the original experiment.
Kelly et al.	Literal Replication	The earlier findings may be reexamined using the same manipulations (independent variables, experimental procedures, etc.) and measures (dependent variables, methods of data analysis, etc.).
	Operational Replication	If the experimenter wishes to vary criterion measures, the experiment would be termed an operational replication. In this instance, the dependent variable would represent a different operationalization of the construct; the essential conceptual meaning would remain unchanged.
	Instrumental Replication	This replication is carried out when the dependent measures are replicated and the experimental manipulations are varied. Variations in the implementation of experimental procedures which do not go beyond the originally established relationship would be included in this category.
	Constructive Replication	A constructive replication attempt may be identified when both manipulations and measures are varied. This replication involves the attempt to achieve equivalent results using an entirely original methods recipe.
La Sorte	Retest Replication	In its general form retest replication is a repeat of an original study with few if any significant changes in the research design. The retest has two major purposes: 1) it acts as a reliability check of the original study, and 2) inconsistencies and errors in procedure and analysis can be uncovered in the repeat. Although the retest increases one's confidence that the findings are not artifactual, it does not eliminate the possibility of error in process, especially when the same investigator conducts both studies.
	Internal Replication	The differences between the retest and internal replication are mainly procedural. Instead of seeking confirmation of an original study, the internal replication is built into the original study design. So the data, part of which are used for the replication, are gathered simultaneously by the same investigator using a common set of research operations. One finds variations in the procedures for selecting the samples. Two of these procedures are: 1) drawing two or more independent samples, and 2) taking a single sample and later dividing it into subsamples for purposes of analysis and comparison. The internal replication provides an additional data supply which acts to cross-check the reliability of the observed relationships. Thus it is methodologically superior to the single study where the hypothesis is tested only once by one body of data.
	Independent Replication	Independent replication is the basic procedure for verifying an empirical generalization. It does this by introducing significant modifications into the original research design in order to answer questions about the empirical generalization that go beyond those of reliability and confirmation. The essential modifications include independent samples drawn from related or different universes by different investigators. These replications differ in design and purpose. They can, however, be broadly categorized into three problem areas. First, is the empirical generalization valid? Second, does further investigation extend it to other social situations or subgroups outside the scope of the original study? Or, third, is the empirical generalization limited by the conditions of particular social situations or specific subgroups?
	Theoretical Replication	It involves the inductive process of examining the feasibility of fitting empirical findings into a general theoretical framework. These replications seek to verify theoretical generalizations. In these replications, empirical variables, which have concrete anchoring points are abstracted and conceptualized to a higher theoretical plane, it is necessary to sample a variety of groups using different indicators of the same concepts.
Leone & Schultz	Experimental Replication	The same experiment is conduced more than once, although there can be (especially with social systems) no perfect replications. It involves the same method and the same situation.
	Nonexperimental Replication	The same method is applied to different situations.
	Corroboration	It involves different method and same situation, or different method and different situation.
Lindsay & Ehrenberg	Close Replication	This replication attempts to keep almost all the know conditions of the study much the same or at least very similar (for example, the population or populations in question, the sampling procedure, the measuring techniques, the background conditions, and the methods of analysis). A close replication is particularly suitable early in a program of research to establish quickly and relatively easily and cheaply whether a new result can be repeated at all.
	Differentiated Replication	It involves deliberate, or at least known, variations in fairly major aspects of the conditions of the study. The aim is to extend the range of conditions under which the result still holds. Exploring a result with deliberate variations in the conditions of observation is the essence of generalization. According to the authors, there are three reasons for running a differentiated replication: 1) Use different methods (different measuring instruments, analysis procedures, experimental setups, and/or investigators) to reach the same result (triangulation), 2) Extended the scope of the results, 3) Define the conditions under which the generalization no longer holds.

Table VI: Descriptions of replications classifications (continued).

Lykken	Literal Replication	This involves exact duplication of the first investigator's sampling procedure, experimental conditions, measuring techniques, and methods of analysis.
	Operational Replication	One strives to duplicate exactly just the sampling and experimental procedures given in the first author's report. The purpose of operational replication is to test whether the investigator's "experimental recipe" the conditions and procedures he considered salient enough to be listed in the "Methods" section of his report will in other hands produce the results that he obtained.
	Constructive Replication	One deliberately avoids imitation of the first author's methods. To obtain an ideal constructive replication, one would provide a competent investigator with nothing more than a clear statement of the empirical "fact" which the first author would claim to have established.
Mittelstaedt & Zorn	Type I	The replicating researcher uses the same data sources, models, proxy variables and statistical methods as the original researcher.
	Type II	The replicating researcher uses the same data sources, but employs different models, proxy variables and/or statistical methods.
	Type III	The replicating researcher uses the same models, proxy variables and statistical methods, but applies them to different data than those used by the original researcher.
	Type IV	In this replication, different models, proxy variables and statistical methods are applied to different data.
Monroe	Simultaneous Replication	Does the same researcher in the same study investigate consumer reactions to more than one product, or to more than one advertisement?
	Sequential Replication	Does the researcher or another researcher repeat the study using the same or different stimuli at another point in time?
	Nonindependent Replication	The replication is conducted by the same researcher.
	Independent Replication	The replication is conducted by different researcher.
	Assumed Replication	For example, a researcher using both males and females simultaneously in a study and finding no gender covariate effect assumes replication across gender.
	Demonstrated Replication	What is preferable is separate gender conditions wherein the effect has or has not been obtained separately for males an females, that is, demonstrated.
	Strict Replication	The replication is a faithful reproduction of the original study.
	Partial Replication	The replication is a faithful reproduction of some aspects of the original study.
	Conceptual Replication	The replication uses a similar conceptual structure but incorporates changes in procedures and independent variables.
Radder	Reproducibility of the material realization of an experiment	In this type of reproduction, the replicator correctly performs all the experimental actions following instructions given by the experimenter who ran the previous experiment. This reproduction is based on a division of labour, where other previously instructed people can run the replication without being acquainted with the theory underlying the experiment. As in this reproduction it is possible to follow the same procedure to verify the outcome without detailed knowledge of the theory, there may be differences in the theoretical interpretations of the experiment.
	Reproducibility of an experiment under a fixed theoretical interpretation	This reproduction implies that the conditions of the previous experiment can be intentionally altered in the replications, provided that the variations are irrelevant to the theoretical interpretation of the experiment.
	Reproducibility of the result of an experiment	This type of reproduction refers to when it is possible to achieve the same result as a previous experiment using different methods. This category excludes a reproduction of the same material operationalization.
Schmidt	Direct Replication	This involves repeating the procedure of a previous experiment. In this replication, the context variables, the dependent variable or subject selection are open to modification.
	Conceptual Replication	This is the use of different methods to retest the hypothesis or result of a previous experiment.
Tsang & Kwan	Checking of Analysis	In this type of replication, the researcher employs exactly the same procedures used in a past study to analyze the latter's data set. Its purpose is to check whether investigators of the original study have committed any errors in the process of analyzing the data.
	Reanalysis of Data	Unlike the checking of analysis, in this type of replication, the researcher uses different procedures to reanalyze the data of a previous study. The aim is to assess whether and how the results are affected by problems of definition, as well as by the particular techniques of analysis. Quite often the replication involves using more powerful statistical thecniques that were not available when the original study was conducted.
	Exact Replication	This is the case where a previous study is repeated on the same population by using basically the same procedures. The objective is to keep the contingent conditions as similar as possible to those of the previous study. The researcher usually uses a different sample of the subjects. The main purpose is to assess whether the findings of a past study are reproducible.
	Conceptual Extension	A conceptual extension involves employing procedures different from those of the original study and drawing a sample from the same population. The differences may lie in the way of measuring constructs, structuring the relationships among constructs, analyzing data, and so forth. In spite of these differences, the replication is based on the same theory as the original study. The findings may lead to a revision of the theory.
	Empirical Generalization	In this replication, a previous study is repeated on different populations. The researcher runs an empirical generalization to test the extent to which the study results can be generalized to other populations. It follows the original experimental procedures as closely as possible.
	Generalization and extension	The researcher employs different research procedures and draws a sample from a different population of subjects. The more imprecise the replication, the greater the benefit to the external validity of the original finding, if its results support the finding. However, if the result fail to support the original finding, it is difficult to tell whether that lack of support stems from the instability of the finding or from the imprecision of the replication.