Reflections on Replications

Fred Niederman  
John Cook School of Business  
Saint Louis University  
niederfa@slu.edu

Sal March  
Owen Graduate School of Management  
Vanderbilt University  
salmarch@owen.vanderbilt.edu

Abstract:
We are immensely pleased to see a publication dedicated to replications in the IS field (Dennis and Valacich, 2014). We believe it is long overdue and will serve a useful and path breaking function in the Information Systems discipline and, perhaps, in the larger academic community (Berthon, et al., 2002). Toward that end, we share our particular view of replication: why it is important and how various types of replications may co-exist. We also offer some thoughts about ways that researchers might go about this type of research that will add value to their own work and more effectively add value to the body of knowledge that represents the IS discipline.

Keywords: Theory, Philosophy of IS Research, Research Methods

The manuscript was received 04/13/2015 and was with the authors 2 months for 2 revisions.
Reflections on Replications

1 Introduction

Replication is a standard practice in the physical science disciplines (Jasny, et al. 2011). If research results cannot be replicated in, for example, physics or chemistry, their validity is called into question (Nature Chemical Biology Editorial 2013). In Popper’s words, “non-reproducible single occurrences are of no significance to science” (Popper 2005, p. 66). The physical sciences have a significant advantage over the social sciences, however, in that they deal with ontologically objective phenomena, that is, with phenomena that are governed by immutable laws of nature—phenomena that exist in “reality” (Bunge, 1977). The challenge in such fields is to discover and codify those laws. Certainly, in such fields, research results must be consistent across research studies, provided the environments in which the studies are conducted are consistent. The expectation is that because the laws of nature do not change, if, in fact, a research study has identified a law of nature (i.e., sufficiently justified theory), then that law should apply across all of reality and certainly across research studies within a given context (McKubre 2008).

Of course, all theories are approximations of natural laws and science advances when contexts are expanded and theories are generalized, resulting in a better understanding of those laws. Newton’s laws of motion, for example, “hold to a good approximation for macroscopic objects under everyday conditions. However, Newton’s laws … are inappropriate for use in certain circumstances, most notably at very small scales, very high speeds … Explanation of these phenomena requires more sophisticated physical theories, including general relativity and quantum field theory” (http://en.wikipedia.org/wiki/Newton%27s_laws_of_motion downloaded February 6, 2015). As scientists develop new approaches and instruments, our understanding of such natural laws and how they apply in a broader array of circumstances is expanded and enhanced. On rare occasions, they are completely replaced (e.g., the Copernican Revolution (Kuhn, 1957)).

Social sciences, on the other hand, have no such ontological foundations. Searle (1995, 2010) argues that “all of institutional [social] reality is created by linguistic representation” (2010, p. 14). It is a “reality” that is created and maintained by human declarations (collective intentionality) for human purposes. Such declarations “change the world by declaring that a state of affairs exists and thus bringing that state of affairs into existence … you make something the case by explicitly saying that it is the case” (2010, p. 12). Examples are as far ranging as: an individual filing a document with an appropriate administrative agency to bring a corporation into existence, to an authorized agent of the government declaring a marriage to exist (“by the powers vested in me … I now declare you man and wife”) to an information system’s users declaring that the system is “useful” or “not useful.” Note that by their nature there is a relative impermanence to such declarations. Corporations dissolve, married couples divorce, and information system’s users may be looking for the “next new thing.” As a result, replication testing of these sorts of “truths” may be accurately conducted but show different findings as the result of contextual change; nuanced interpretation may conclude that samples have been drawn from different populations (e.g. “new” and “old” users may be representative of different populations by virtue of technological or societal changes).

Thus we argue that the “laws” at which social science theories aim to address are themselves most frequently constructions and deconstructions of human intention. Such theories are perishable and context dependent in that the humans, their circumstances, and their intentions can all change. However,

1 This paper is predominantly framed from the perspective of the natural/physical sciences as applied to social science. We recognize that other perspectives are possible and do not intend to imply that replication studies should be limited to this one paradigm.

2 Searle differentiates “brute facts” that are ontologically objective, existing independently from human intention or knowledge, from “institutional facts” that are epistemologically objective, depending for their existence on collective acceptance. We argue that it is epistemology and not ontology that forms the basis of theorizing in social sciences.
the declarations or “constitutive rules” underlying societies (Searle, 2010), while dependent upon human acceptance, are relatively stable (Epstein 1980). For example, as March and Allen (2014) point out, “it is solely by collective intentionality that private property ‘exists.’ The physical object exists independent of human intentionality or even of human knowledge of its existence (Bunge, 1977). However, its status as private property requires human intentionality and the imposition of a status function recognized within the social context. In Bunge’s terminology, private property is a “fiction” that exists solely in human minds” (p. 3). Yet we routinely accept that property rights will be honored and make that assumption in social science theories related to incentive structures and compensation. However, it must be recognized that constructs such as private property rights can, and have been, subverted by governments, acquisitive institutions, and unscrupulous individuals in numerous social contexts.

We consider the information systems discipline to be a social science, as opposed to a physical science. Although concerned with the “IT artifact” (i.e., software executing on hardware and utilizing data), which has certain physical characteristics related to space and time, an “information system” is an idea, a conceptualization, a set of procedures and rules that drive the physical components (i.e., electrons moving through circuits and magnetic charges sitting on storage devices as well as the people who interact with them). The information systems literature is concerned not primarily about the physical properties of the IT artifact, but more prominently about the situated IT artifact as utilized and appropriated by human intentionality. Hence, the theories developed and tested must recognize the social context in which an IT artifact is situated. Replication of a study in a different context, thus provides evidence for the social generalization of extant theories and/or identifies social contexts in which those theories are deficient. A replication ultimately can address concerns such as: changes in technology affordances over time, changes in how technology is used and integrated in users’ environments, changes in users themselves that result from enhanced technology awareness, availability, and use, changes in user purpose and intention, and changes in the judgments users make relative to outcomes and effects of knowing about, having, and/or using information technologies (e.g., individual and organizational learning).

Specifically in information systems research, we argue that virtually all environments in which research studies are conducted represent (somewhat) different contexts. For example, studies that use undergraduate students in a laboratory setting at one point in time are likely to represent a different context from the exact same study in exactly the same laboratory with a different set of undergraduate student subjects (or perhaps even the same set of undergraduate students at a different point in time). We do not see this necessarily as “measurement error” or a failure of “test-retest reliability” but as a potentially accurate demonstration of shifting intentions, preferences, or learning among subjects. The shift of preference from Facebook to Instagram among some generations of users, for example, may represent a valid change in the Internet environment rather than a reliability failure of measurement instruments. We argue that differences in the evolution of technological sophistication and understanding of IS concepts over time could have significant impacts on research results (one might argue that this evolution is particularly significant among undergraduate students in their role as technology users). Thus we argue for sensitivity to the meaning of findings that do not support prior research outcomes. On one hand, these may signal methodological deficiencies, invalid theoretical statements, or statements that have been generalized beyond their appropriate context. On the other hand, they may indicate a meaningful change in collective societal preferences or judgments.

That is not to say that we endorse unnecessarily repeating the same study with the same population ad nauseam, but that we endorse identifying important characteristics of the research study that could potentially impact the results as related to the theory in question. We argue, with March and Smith (1995, p. 256), that “Theorizing in IT research must explicate those characteristics of the IT artifact operating in its environment that make it unique to IT and require unique explanations. Theorizing that Newton’s theory of gravity holds for IT, and testing it by dropping a PC from an office window in the MIS department is obviously not valuable.” While extreme, the example is a stern warning that developing a replication study requires significant attention to those characteristics that make the replication study valuable. If the theory addressed in a replication study is not well supported in the literature then a study that directly replicates

---

3 Bunge uses the term “fiction” not in a pejorative sense or in any way diminish the importance of such concepts in human society, but only to distinguish concrete objects from conceptual objects, specifying that the domain of ontology is concrete objects. Unfortunately, others may not interpret the intention so neutrally and be led to ignore or dismiss socially created “facts” as “mere” fiction.

4 It could also indicate a broadening of our observation of a population that is difficult to fully capture from a single sample. Note that even in a small sample, a central tendency may be identified more readily than the full range of possible states including exceptions and outliers within the full population.
Reflections on Replications

Volume 1
Paper 7

prior research as closely as possible is valuable theory testing. However, blindly replicating a series of studies that have adequately justified an extant theory is less so.

We note that replication is becoming an important topic in other business research areas (Hubbard, et al. 1998), exemplified by the Strategic Management Journal special issue on replication (currently in process, see the call for papers at http://smj.strategicmanagement.net/pdf/SMJ_Call_for_Papers_Replication_Special_Issue_7_12_13.pdf, downloaded July 9, 2015). We acknowledge the excellent papers by Dennis and Valacich (2014), Berthon, et al. (2002) Easley, et al. (2000) and Tsang, and Kwan (1999) that each adeptly motivate the need for replication in the social science fields. Dennis and Valacich (2014) further identify three major types of replications: Exact, Methodological and Conceptual. They conclude, “All replications are valuable in advancing science. Conceptual replications are the strongest form of replication because they ensure that there is nothing idiosyncratic about the wording of items, the execution of treatments, or the culture of the original context that would limit the research conclusions” (p. 3).

As discussed above, we argue that, to some degree, by its very nature, replications in the information systems discipline, occurring at different points in time, would have at least some element of conceptual replication, with little hope for exact replication, even in reasonably controlled environments (depending on the length of time and the relevance of technological advances between the studies). Easley, et al. (2000), among others, make a similar argument for experimental studies in all social science contexts, “because uncontrolled extraneous factors have the potential to interact with the various components in an experimental setting” (p. 85).

We further argue that the context for IS research is both technological and social. As discussed above, the technological landscape is ever-changing. Following Searle (2010), we argue that the social landscape is constructed by human intentions and, although relatively stable, it too is subject to change. Thus, replications in the most realistic environments possible are likely to have the most significant impact on the ability to develop theories that are broadly applicable and practical (Van de Ven, 1989), even if perishable.

In this paper we intend to add to that body of knowledge by addressing the following questions:

1. Why are replications important? What should motivate replication studies?
2. How should we interpret the results of replication studies?
3. How can different methods be used to support replication studies?
4. How might replication be used with different types of meta-theories?
5. What is the minimum content we should expect from a replication study?
6. What is an example of how an existing study might be used for replication?
7. What differentiates replications in IS from replications in other fields?

We conclude with a discussion of a conceptual replication study (Allen and March 2012) and the resultant dialogue in the research literature (Shanks and Weber 2012). We expect that replications may elicit many citations (i.e., where authors doing work in the same area will want to be comprehensive in how they address what is known on the topic), but more importantly that they will lead to a robustness in identifying and explicating general principles (or meta-theories) that transcend specific technologies.

We do not intend our discussion to be interpreted as a “required recipe” for conducting replication studies in the IS field. Rather, we provide a set of constructive, suggestive, and helpful guidelines, the overall spirit of which is more important for any replication study to grasp than the minutia of any step-by-step details.

2 Key Questions about Replication

2.1 Why Are Replications Important? What Should Motivate Replication Studies?

If we are interested in establishing a field in which knowledge is accumulated and theories are justified, we need mechanisms to re-produce, verify, challenge, and extend reported research results (see, e.g., Nature Publishing Group, 2015). If we do not replicate extant studies, we are condemned to having very
little support for the posed theories and reported results and we open the discipline to “one off” or even fabricated results (see, e.g., Fanelli 2009 and Silverman 2014). Although we recognize that a theory cannot be ultimately proven (Kuhn, 1962), particularly in the social sciences, we argue that with sufficient replication we can gain confidence that it is useful, at least within a specific context. We further argue that differentiating replication studies from theory testing studies is frequently semantic rather than substantive and urge authors, reviewers and editors to recognize the importance of gathering both confirmatory and falsifying evidence for extant theories.

Furthermore, there are many definitions of “theory.” At root, most definitions revolve around theory as a set of statements specifying relationships among entities of one sort or another that can be falsified (Popper, 1963) or tested (Burton-Jones, et al., 2014). While such a definition of theory is generally abstract, it can be applied to particular instances for prediction and/or explanation (Weber, 2012). As such statements hold up to or are refined (and sometimes discarded) by repeated testing, our confidence in them grows (or diminishes). As noted above they are never “proven” because the next instance may react to different emergent phenomena in the world, as demonstrated by the “black swan” phenomenon (see, e.g., Taleb, 2010). However, with repeated testing we can grow a degree of trust that by following the predictions and explanations of a theory we can act with a higher probability of success, particularly within the range where the theory is demonstrated to be applicable. Because of its ability to be tested and reshaped to more accurately describe the patterns of findings in the world (including the world of ontologically-subjective social reality (Searle, 2010)), theory serves as a very powerful tool for the accumulation of knowledge. As theory is proposed and tested, the resulting articulation either becomes more nuanced and gathers more support, forming a basis for rigorous codification of knowledge and for application to new situations both for further testing and for use in practice or, alternatively, may generate a variety of overlapping or conflicting theories offering the opportunity for further study into why such diverse results are observed.

In this view of theory, testing and re-testing are the key links to progress in the accumulation of knowledge within a field of study\(^5\). Note that by this definition, even from its first proposal without any supporting evidence, a set of statements that can be falsified should be considered “a theory.” However, a theory with no support does not (and should not) generate much confidence. How would we, therefore, gain confidence regarding the “truth” or “utility” of these statements? The method is through testing and re-testing (replication). Replication in its most broad definition is the mechanism for gaining confidence in a theory. We recognize that people of good will may perceive the range of valid replications broadly or narrowly. We tend to view it broadly and will explain in the following discussion both what this consists of and why we think such a broad view is helpful.

We see theory as having a lifecycle. Our preferred metaphor is that of a seed that transforms to a seedling then, eventually, into an old growth tree. Whether the origin of a theory is inductive or deductive, it must begin somewhere. If it originates inductively, say through a grounded theory approach, it is largely untested at its inception. Granted that the grounded theory process, particularly if theoretical sampling is used, may be viewed as a kind of testing where each data collection event serves as a basis for testing the emerging theory. This at most represents a limited range of tests. Even with the purest of intentions, theory tested by those who construct it must be considered tentative relative to theory tested and verified with others independent from its originators.

Hence, we answer the question of importance and motivation with a single response: it is crucial to the nature of scientific inquiry and it is our duty as serious researchers to build a cumulative knowledge base. These can be accomplished most effectively through replication studies that scrutinize existing results and lead to an open and transparent discipline.

\(^5\) In principle a field will not have much leverage in the world of academia if it does not have useful theories to champion. At first glance this would seem to mean that a field where theories are proposed and then shown wanting through replication would be self-destructive. However, we would argue that having untested and weak theory may be little better than having none. The way to strengthen and build theory is more likely through a vigorous program of proposing AND testing -- with the proviso that the testing is not aimed at simple rejection, but rather at proposing new and stronger versions (e.g. data gathered from finance industry may not work in manufacturing, but such testing may suggest contingency rather than outright rejection of the theory as it applies in financial industry settings).
2.2 How Should We Interpret The Results Of Replication Studies?

We expect that results from replications will fall broadly into three categories: they will support the theory underlying the original study, they will refute it fully, or they will support it in part and refute it in part (Dennis and Valacich, 2014). What conclusions are likely as a result of these differing types of results?

Where there is total support, the replication authors will be able to acknowledge the increased confidence with which the theory should subsequently be viewed. It may be difficult to quantify such change in confidence. We simply note that (1) as we gain confidence in a theory through repeated exact replication, the importance of additional exact replication studies diminishes; (2) assuming equally rigorous research methodology and confirmatory results, replications by teams outside the sphere of the originators of a theory should have a greater impact due to both the necessary clarity of the original work in allowing replication and the assumption of “disinterest” of the outside party (conversely for dis-confirmatory results); and (3) again assuming equally rigorous research methodology and confirmatory results, where the replication stretches the scope of the theory (conceptual replication) we argue that such studies increase our confidence that the theory is more generally applicable and comes closer to the notion of a “law” of social science. Furthermore, we urge researchers engaged in replications to clearly identify and articulate the differences between the original study and the replication, even if the intention is to develop an exact replication, at least some differences in context are inevitable (as discussed above).

In the second case, where a replication refutes fully a theory, we must carefully consider a range of explanations for the divergence. It is possible that, for example, (1) the original theory is simply incorrect or misstated and should be abandoned; (2) the initial or replication samples, even if random, were drawn from a non-representative collection of instances of the population; (3) that the samples come from different populations – that they are on different sides of a boundary which may relate to different demographic groups, different stages of technology maturity, or just the passage of time; (4) that there were, in fact, conceptual or measurement errors in one or the other study; or (5) that the replicating authors misinterpreted either the content or procedures of the original (note, however, that partial “blame” for this would also accrue to the authors of the original study for failure to unambiguously describe their study). This is clearly not a comprehensive list, but it shows the complexity of interpretation that will face the replicating authors (and reviewers of such submitted papers). In any case, rarely will one replication with contrasting results suffice to dismiss a theory, but rather will likely call for additional study to resolve such differences (Diesing, 1991, p. 33 and p. 251).

For example, “perceived ease of use” of a technology (Davis, 1986, Venkatesh and Davis, 2000) is likely to have a very different meaning for IS users today than it had 10 years ago. Replicating a study using 10-year old technology is likely to produce different levels of “perceived ease of use” for the same technology that was used in the original study, but perhaps the same or higher level of “intention to use” because the technology has become institutionalized or extended to a broader array of applications. Thus, a replication may yield the results: “compared to my iPad and Google the 10-year old technology is not easy to use,” but “it is the best available in my context so I intend to use it.” That should not be taken as evidence that “perceived ease of use” is not an important predictor of “intention to use,” but rather that perceptions of “ease of use” change over time relative to particular technologies. It also changes relative to technological advances, where an application that is easy to use with a large screen and full keyboard may be difficult to use on a mobile device without significant change in the user interface. Perhaps there is a “hurdle level” of ease of use after which increasing ease of use may not impact “intention to use” in certain organizational and social contexts. Similarly, concerns over privacy and security may mitigate intention to use even for the perceived easiest to use and most useful technologies.

2.3 How Can Different Methods Be Used To Support Replication Studies?

Of course, the most likely type of replication will use the same methodology as was used in the original study. Thus an experiment will likely be used to test the results of a published experiment; a survey for a survey and so forth. Experiments that attempt an exact replication (Dennis and Valacich, 2014) of other experiments would ideally try to reconstruct the design, the method of assigning subjects to treatment and control conditions, and the design of procedures for treatment and control. How closely does a replication have to adhere to the original study? We would anticipate a wide range of studies with the standard being literal copying of all elements of the study. In an experimental setting this would entail providing as exactly as possible the same treatment and control conditions; using as closely as possible the same laboratory setting; and selecting as closely as possible the same subject pool. Where the experiment is
based on student subjects slight differences such as the major or concentration of students, their ages, their distribution by gender, and others may naturally vary as a function of the near impossibility (and small economic payback) of gathering an identical copy of the original. Similar logic can be applied to surveys. One would expect the use of the same instruments, perhaps with mild editing for differences in specific technologies, distributed in roughly the same way using a similar sampling approach of the same population (or as nearly the same as technologically possible). Gathered data would be analyzed using the same mathematical techniques with the same tests for meeting distribution and other statistical prerequisites. We would expect that advances in the range and subtlety of analysis techniques would be presented as alternative calculations, in some cases showing that different techniques largely produce similar results or that more sophisticated techniques can provide additional insights.

Perhaps more interesting will be studies that use a different method from that used by the original researchers. Using an experiment, for example, to replicate findings derived from a case study or survey would require significant translation of the constructs in the original into a new set of measures and into procedures that would realistically mirror those of the original. Ease of use, for example, might be clearly understood by survey takers, but may present an interesting challenge for researchers to instantiate in an experimental setting. We might need a sequence of first measuring alternative technology configurations and measuring speed, accuracy, and perceptions of ease of use to establish a baseline for particular technologies, before testing these against other variables in the experimental setting. Additionally, ease of use may vary for the same technology across a range of tasks raising the question of whether the concept of ease of use really ought to be viewed as an attribute of a particular technology or as an interaction of the technology and user for particular tasks. In such a way, however, concepts like the effect of ease of use on actual use or on adoption and adaptation of new technologies might be examined more completely (Goodhue, et al., 2000). Findings from such experiments might prompt a new generation of more nuanced survey questions to quantify and refine the dynamics by which ease of use can be built into new artifacts.

Case studies present a rich and open set of possibilities for replication. We see the notion of replication in the use of multiple case studies where, when analyzed sequentially, subsequent case studies can act as a kind of natural replication of the original. By the same logic, case researchers can look for instances of varying similarity and replicate as best as possible the methodologies of earlier researchers. It is completely logical that a case study can generate theory, particularly as defined by a set of statements with varied levels of support and confidence. As such theory is made explicit, statements may be subjected to testing. Yin (2013) specifies at least two approaches to such testing – taking cases where a phenomenon is expected or other cases where it is not expected and support or refute theory when such phenomenon are (or are not) so observed. By this logic, the results of earlier case studies can be tested in new situations where they are expected or not expected to apply.

Design Science and action research would also seem to be especially rich areas for the creation of replication studies. Ideally, design science and action research studies make two contributions: they create and explain the value of the features or attributes that add benefit to a particular IT artifact (or class of IT artifacts) and they create and explain the procedures or processes that led to better (or worse) design or practice outcomes. Scholars should be able to take the findings of design science or action research studies and examine or test the lessons learned in new situations that are more or less similar. For example Schmidt-Rauch and Schwabe (2011) proposed characteristics that aid in the design of “recommender systems” using the case of travel agents. Perhaps similar characteristics used in the design of recommender systems for helping clients select insurance plans or investment portfolios would be aided by applying the same techniques. Schmidt-Rauch and Schwabe (2011) also described a pattern of building, testing, and then modification of such systems. Designers and action researchers in any software domain should be able to observe the effectiveness (or lack thereof) of using this sort of approach. Perhaps it works well, but is of too high a cost outside the research environment. This would be a worthwhile contribution for the field to have available. Design science and action research can also attempt to replicate findings from other sources of theory. For example, they provide a natural mechanism to address questions such as: Do better designs come from attempting to build artifacts that are “easier to use” or “more useful”? Do we see better products if we build them around a task-technology fit paradigm?

Grounded Theory has generally been applied to understudied phenomenon with the intention of producing new theory. It is clear that a fair amount of interpretation is applied to the collected data, typically derived from a set of interviews conducted pertaining to a particular phenomenon. Within the grounded theory method through theoretical sampling there is a certain amount of within-study replication (e.g., from one
interview to the next) and a certain amount of theory testing by extending the range of interviewees along one or more dimensions. It is not clear why independent researchers cannot profitably return to a similar population, replicate the interview protocol, and develop alternative coding schemes if so led by the actual data. In classroom grounded theory exercises, we observed significant variance in interpretation from one student to another regarding the same transcript content. Another research group using the same data (or collecting new data) may arrive at alternative theory and explanations that shed new light on the same phenomenon (or reinforce the earlier findings.) It would seem there would also be significant power in testing theory built from other methods using grounded theory. For example, in the classroom setting, we used an interview approach to gather data on adoption and use of a new technology. While we found some support for the constructs of TAM (Davis, 1986), we found strong alternatives pertaining to cost, ownership, flexibility, applicability to particular problems, and aesthetics. It was our experience that, in principle, there was sufficient reason to apply a grounded theory approach as a means of replicating or testing prevailing ideas built through other research approaches.

2.4 How Might Replication Be Used With Different Types Of Meta-Theory?

Burton-Jones et al. (2014) argue that three research perspectives describe the types of concepts and relationships used to construct theory: process, variance, and systems. While variance theories are the most common in the IS discipline (Rivard, 2014), a variety of methods can be used in each perspective. Variance theories generally focus on co-variation among properties, process theories focus on sequences of events and systems theories take a holistic view of phenomena, focusing on their interacting parts and emergent properties.

In a variance theory perspective constructs are viewed as being in particular relationships. Researchers replicating variance theory studies may choose to examine the particular constructs and relationships exactly as described in the original studies. However, based on data collected, examination of alternate models may provide additional insights and opportunities to test the shape and construction of models as well as the relationships among constructs (e.g. reversing the direction of a relationship, observation of independence of mediating or moderating variables).

Process theory provides a very rich set of possibilities for replication. We see process theory originating from observations that a particular set of actions in a particular order can be associated with particular outcomes. Once such a process is defined, replications can look at the ability to follow the same process in different settings (e.g. using a sequence of risk management steps developed in a project management setting for estimation of risk in an assembly line setting) and addressing issues such as: (1) the degree to which adaptations or exceptions are necessary to make the process useful in the alternative arena (2) the degree to which following the posed process drains creativity and license to innovate (e.g., versus saving time and avoiding “reinventing the wheel” phenomena) and (3) the effects of variations in activity content, intervals between activities, and the particular sequence of activities on ultimate outcomes. For example, a process defined as a sequence of steps in developing a risk management practice for IS projects (identifying risks, listing potential actions to avoid these, funding all or some of these actions, taking the actions, and evaluating project outcomes) may look at variations in sequencing and implementation in terms of methods for identifying risks, decision rules for including and excluding possible risks, using bottom up or top down approaches to funding projects, altering actions (and timing) of efforts toward ameliorating risks, and different methods (and timing) for assessing outcomes in additional IS projects as a way of establishing and refining process theory pertaining to risk management in IS projects.

Systems theory also provides a rich set of opportunities for replications. We see systems theory as comprised of examining the effects of fixed systems (e.g., the technologies that comprise part of an extant information system) and the effects of open systems (e.g. where inputs are somewhat unpredictable). To study fixed systems, we would look at the variation of reactions to the introduction of such systems in different environments. We would see, for example, studies of the effects of introducing say ERP or cloud-based systems being open to replication and testing to determine if the observed results are consistent across variations in inputs, purposes, funding, personnel, and other system attributes.

Furthermore, replications could contrast variations in system designs, ideally using principles derived from the original studies to determine the extent of their generalizability. The search for “best” organizing principles would suggest many replications of various systems and rich data gathering and analysis to identify both the quality of such principles and the methods for best utilizing them.
As a general viewpoint, we see replication being useful for any meta-theory as well as for any research perspective. We would see issues of theory level – grand, mid-range, or substantive (detailed), to be open to replication. For example, we would see a relatively "grand" theory such as task-technology alignment tested in grand, mid-range, or substantive levels looking at a more restricted scope but perhaps more detailed sets of constructs.

2.5 What Is The Minimum Content We Should Expect from a Replication?

It would stand to reason that a replication study must report (1) the theory, context and methodology of the original study or studies that are being replicated; (2) the procedures used in the replication; and (3) the comparative results of the new study. Regarding the original study on which the replication is based, clearly the replication authors must summarize enough of the original for a thorough understanding by readers but not so much that the reader is overwhelmed by details best left to those who wish to track down the original research. At a minimum, the summary of the original study should provide a flavor of the research setting, original method in broad brush (with details to follow in the method of replication section), theory basis (if there is any provided), and results (Jasny 2011, Nature Chemical Biology Editorial 2013, Nature Publishing Group 2015).

The reporting of the procedures used in the replication should indicate where there is an attempt to exactly replicate the original study and where there are variations, intended or resulting from unintended circumstances. Where there are variations, the researchers should explain the nature and purpose of the variations. Variations may, for example, be selected due to (1) evolution of technologies or processes since the original study; (2) updates to standard research methodology practices such as moving from lesser to more sophisticated statistical tests; (3) lack of access to equivalent respondents, laboratories, or other resources (and hence the need for substitution); (4) a deliberate decision to test theory in a different environment or (5) opportunities that include different attributes (e.g. another case site that mirrors inexacty the original or extends near or beyond the original theory scope). The readers should have enough information to assess the degree to which the replication mirrors and deviates from the processes of the original. It is our contention that even in survey and experimental settings, replications will rarely, if ever, be totally exact, so the question is how to convey to the reader where there is deviation and how much deviation there is. Clearly the more deviation the more these differences themselves can suggest an explanation for varied findings. However, over time, some types of deviation may prove more influential than others.

In addition to simply providing the results of the replication’s analyzed data, we contend that it will be crucial to contrast and compare the new findings with those of the original study or studies. We have discussed in a prior section the general view that these findings may totally support or refute the original; however, they may also partially support the original. Moreover, we expect a variety of types of explanation (e.g. flaws in original theory, variations in scope, influences of differences in method, etc.) and consideration of these possibilities should be explicit.

We would see these requirements applying in a non-identical, but equivalent manner, for replications using different methods and across levels of theory (grand, mid-range, and substantive or detailed). In each case examination of the original work, the process of the replication, and the replication findings in the context of original findings would be important information to report (see, e.g., Allen and March 2012 and Adams, et al., 1992).

2.6 What Would Be an Example of How an Existing Study Might Be Used for Replication?

In order to illustrate our thinking regarding replications, we next discuss a particular study that we feel illustrates how replication can proceed. Holmström and Sawyer (2011) examine impediments toward optimal requirements gathering by interpreting them within a social situation. The study is based on interview data with 26 system developers from 5 small IT consulting companies. This is supplemented by more than 200 documents processed using a grounded theory three step coding approach. The authors formulate a conclusion that too often developers simplify rather than document the full richness of requirements for at least two reasons: (1) the range of participating actors from whom requirements are solicited are too homogeneous and (2) methods focus on actionable rather than the full range of requirements. As a result the authors conclude that requirements are not "things" as we would think of a
tree with existence beyond our labeling of it; but rather as social constructions resulting from discussion, negotiation and conflict resolution.

In our view these observations suggest (at least) two questions worthy of replication and perhaps elaboration. First, does the level of detail and completeness of requirements gathering change with the number or diversity of actors providing input? Second, does the use of alternative methods for requirements gathering affect the level of detail and completeness of requirements gathering?

We would call such a follow-up study focusing on these questions a replication of the Holmström and Sawyer (2011) study. At a first pass, such a replication study could also use a grounded theory method, interviewing developers about their practices in requirements gathering, focused on actors, methods, and results. Using such an approach might provide supporting, refuting, or extending observations, the reporting of which would increase or diminish confidence that we might have in the robustness of the original findings. Alternately, a replication study could shift methods and test the frequency and level of agreement regarding these relationships through crafting a survey and distributing it widely. To the extent that activities can be developed and administered within the scope of the resource limitations, experiments could also test the degree to which more thorough requirements can be derived from more diverse sets of actors. Perhaps such experiments might also suggest mechanics of discussion, negotiation, and conflict resolution that lead to fuller and richer requirement sets.

In the end we see a powerful role for replications of this type (and others) in moving forward our understanding of the requirements determination process. We see these as theory testing, building, and extending opportunities. Note that we selected this particular study because of the centrality of development of new systems (whether they be classical functional area systems like those supporting HR or marketing or innovative systems such as mobile apps and extensions to social media sites) and the importance of requirements gathering in that process. Note also that in addition to theory, such studies can provide insights into practice through critical success factors or definition of processes linked to achieving better outcomes.

2.7 What Differentiates Replications in IS From Replications in Other Fields?

As discussed above, although there are general principles for conducting replication research, the IS discipline presents a number of challenges beyond those inherent in social science research in general. The evolving technological context in which IS research is situated presents the most obvious challenge. As technologies change, both humans and organizational contexts adapt to those new technologies. However, it is human and organizational purposes and constraints that drive the changes in technologies. Theories dependent upon human, organizational or technological limitations are apt to have a short shelf life and replications are apt to be moot.

Furthermore, given the current emphasis on theory, particularly in most if not all of the senior scholars’ basket of 8 journals, IS researchers are incentivized to “borrow” theories from reference disciplines and to view IS phenomena through such theories, i.e., as a “theoretical lens,” rather than developing and testing indigenous IS theories. We are concerned that the use of such reference discipline theories could emphasize issues that are outside the IS context while failing to capture the unique interplay of influences that characterize the IS phenomena (see, e.g., Orlikowski and Iacono 2001, Weber 2003, Orlikowski and Iacono 2006). If observation of users engaged with technology, for example, mirrors people’s cognitive actions while doing other activities, we fear that conclusions about user behavior may be shallow relative to the richness of behavior when the user, the technology, the task, and goals/objectives may all influence one another.

We believe, therefore, that replication by IS scholars of indigenous IS theories will bring an especially positive virtuous cycle of growth and knowledge accumulation to our field. We would especially like to see theories such as Goodhue and Thompson’s view of task-technology fit (1995) and DeLone and McLean’s success model (1992) explicitly used to test and refine more specific studies within these constellations. We also argue that replication studies utilizing socio-technical theory and socio-materiality theory would be promising for enriching our cumulative knowledge of IS in a meaningful way. Further we see a role for replicating atheoretical studies or studies in which the theory is more implicit than named and delineated (e.g., Holmström and Sawyer, 2011). In such studies the goal of replication would be to formalize the results into indigenous theory within the IS domain. We argue that topics such as requirements determination and information system implementation, use and management hold particular promise for the development and testing of indigenous IS theories (acknowledging that we share many of these
concerns with other fields such as architecture, engineering, organizational design and new product development, however, asserting that the intangibility of the IS context presents differing and unique conceptualizations). That said, we see a number of challenges to replications in IS. As discussed above, technology changes so rapidly that particular features and affordances may come and go before their testing has run its cycle from start to end. We saw in the 1980’s, for example, a large body of work on tables versus graphs and decision maker response to these representations (see, e.g., Vogel, et al., 1986 and Taylor and Wang, 1987). This literature has been largely abandoned, presumably because we can now change tables to graphs or vice versa with a few keystrokes in Excel or other BI tools and there is no reason any decision maker cannot make use of both with virtually no cost. Other challenges include the fact that many interactions in IS phenomena involve mental activity – learning, decision making, prioritizing, evaluating, that are: (1) fluid even during short periods of time, (2) can occur through different activity pathways, and (3) are extremely difficult to document. Finally, gaining access to field environments and those actually performing IS professional tasks can be difficult, time consuming, and risky, in the sense that once data is gathered, it may or may not address the originally posed questions.

3 Discussion and Conclusions

In this essay we have offered our view of replication: why it is important, how various types of replications may co-exist, and offer some thoughts about ways that researchers might go about this type of research that will add value to their own work and more effectively add to the body of knowledge that represents the IS discipline. We agree that replication is the foundation of scientific knowledge (Berthon, et al., 2002), but argue that performing replication studies in the IS discipline is fraught with challenges. Some of these are self-imposed by reviewers and journal editors (see, e.g., Zmud, 1996 and Lee, 2010). Some are due to the nature of social sciences, in general, and others are due to the nature of information systems phenomena.

Knowledge is advanced in the natural sciences when theory is developed to explain heretofore unexplained phenomena or when a context is identified in which extant theory, previously justified and accepted, is demonstrated not to be so (Popper 1963). We urge researchers in the IS discipline to take a similar approach: identify heretofore unexplained phenomena or contexts in which there is significant potential that extant theory is inadequate. The crucial element is demonstrating that there are serious reasons to believe that the replication will add to knowledge: that the results of prior studies are not due to chance or to unidentified factors or that the context is significantly different from those of prior studies. In either case a replication that supports the theory has significant value as does a research study that does not support the theory. Given the broad range of differences between IS contexts and contexts in which extant social science theories have been developed we believe there are significant opportunities for contributions both to the IS discipline and to reference disciplines from which theories are borrowed.

The design science paradigm (Hevner, et al., 2004) has become more common in IS research studies, focused on addressing problems in various IS contexts (see, e.g., VanderMeer, et al., 2014 and Meyer, et al., 2014). These types of studies are likely to become even more prevalent with the explosion of text and data mining as mechanisms to explore and draw conclusions from data without deriving abstract conceptualizations (see, e.g., Shmueli and Koppis, 2011). We posit that in many of these studies there is implicit or emergent theory (at least with respect to the performance measures used to evaluate IS artifacts in situ). If we view theory as a set of statements and relationships among them, we can infer these statements from the nature of the data collected and their findings. For example, a design science study showing that having more communication channels is strongly associated with more participation in a virtual team setting, will imply a theoretical statement about the relationship between channels and participation. Note that such a statement can take varied formats and can have very different meanings based on small variations in wording. The statements, “more channels causes more participation” is very different from “more channels can cause more participation.” We would suggest that such wording is important and that a progression of wording from “loose” (e.g. it “can” cause a result) to “medium” (e.g. it “usually” causes a result) to “deterministic” (e.g. it “always” causes a result) can be one of the benefits of replication – more tentative language can evolve toward more certain language as testing progresses and confidence is gained. Conversely, even if a study shows that an effect is not “always” caused by a precursor, it may provide strong evidence that it frequently does.
Other studies borrow a “theoretical lens,” generally from outside the IS domain, to provide terminology and perhaps relationships, measures, and techniques for examining research questions pertaining to an IS domain, e.g., the set of conceptual modeling studies based on Bunge’s (1977) ontology (e.g., Bodart, et al., 2001, Gemino and Wand, 2005, Bowen, et al., 2006, Bowen et al., 2009, Shanks, et al., 2008, Allen and March, 2012, Shanks and Weber, 2012). These sorts of studies present some interesting challenges for replication. If the target of interest is the original theory or other conceptualization which is being used as a theoretical lens, then the replication would address the degree to which such theory “holds” in the IS context. On the other hand, if the target of interest is the IS domain, adherence to concepts of the theoretical lens can be loosened.

Consider, for example, the conceptual replication study performed by Allen and March (2012). This study presents dis-confirmatory results to those presented by Shanks, et al. (2008). The Shanks, et al. (2008) study uses the theory of ontological clarity as posed by Wand and Weber (1993) and further articulated by Burton-Jones, Wand and Weber (2009) to study the “whole-part” relationship construct in conceptual modeling. The theory predicts that users’ understanding of a conceptual model will be undermined whenever it lacks “ontological clarity.” The study operationalizes “ontological clarity” using the relationship construct in an experimental setting (Shanks, et al. 2008):

When relationships are used to represent composites, construct overload arises, because a single modeling construct is used to represent two ontological constructs (a mutual property and a composite). As a result, ontological clarity is undermined, and Wand and Weber (1993) predict that users of a conceptual model will have greater difficulty understanding the semantics of the real-world domain represented by the model (p. 557).

The results of the experiments are concluded to support the theory. While Allen and March (2012) raise methodological concerns, they also execute a conceptual replication yielding dis-confirmatory experimental results. They conclude:

Given that we have provided compelling evidence that construct overload does not result in inferior human performance in the use of conceptual models, the theory of ontological clarity has been falsified (Popper 1963) and must either be discarded or modified to account for the new experimental findings (p. 962).

We believe that the dialogue represented by this series of papers is a healthy and appropriate exchange in conceptual modeling research (Shanks and Weber (2012) is a response to the Allen and March (2012) study). While we do not believe a single dis-confirmatory study is sufficient to falsify a theory, we believe it is sufficient to bring the theory into question. Furthermore, interpreting the results of the Allen and March (2012) study is not straightforward, as argued by Shanks and Weber (2012). The exchange raises significant methodological and conceptual questions that, hopefully, will lead to additional research in the area, perhaps using different research methods or different meta-theories. Other exemplars include Samaddar and Kadiyala (2006), Kettinger and Lee (1999) and Adams, et al. (1992).

In our capacity as reviewers and members of editorial boards, we see a number of papers that (1) add a variable or two to a nomological net or (2) take variables from multiple sources and recombine them for a particular study. In the former case, we argue that an analysis of the nomological net with and without the added variables can constitute a “replication plus.” In the latter case it is more difficult to see this as a replication. The latter does raise the interesting view of research as building theory about particular domains from a collection of variables in particular categories and applying those most logical to the particular domain at hand. For example, the DeLone and McLean model of IS success (1992, 2003) provides a wide range of variables organized into larger units which researchers can select based on some analysis as most likely to explain a particular set of relationships. This has the benefit of allowing flexibility for particular issues, a broad potential for developing standardized and tested instruments, and the ability to “roll up” findings from individual studies into statements about the more abstract formulation based on the overall model. Perhaps the notion of replication in a broad sense can be used to describe these types of studies – even though applied, for example, to a very particular IS context, the selection of standard variables from a larger theory (e.g. DeLone and McLean, 1992) can be viewed as a sophisticated replication. Of course the authors may be more interested in the understanding of the domain, but in passing make an additional contribution to larger theory.

Rather than argue about whether or not a study using the theory and technique of one study but applied to a different population or subset of a population is a “true” replication, we would offer that the value of such studies is potentially high and that defining these “in” rather than “out” of replications has merit.
References


About the Authors

Fred Niederman serves as the Shaughnessy Endowed Professor of MIS at Saint Louis University. His PhD is from the University of Minnesota in 1990. His research interests include IT project management, global information management, MIS personnel, and using MIS to support teams and groups. Recently he has been investigating the integration of MIS functions after corporate mergers and acquisitions. He is a proponent of grounded theory and theory building as a way to enrich the MIS discipline and build intellectual content customized specifically to our field of practice. He has published more than one hundred articles in leading research journals and refereed conference proceedings. He serves on editorial boards for Project Management Journal, TMIS, JAIS, CAIS, Human Resource Management, Journal of International Management, and IEEE Transactions on Engineering Management. He has edited or co-edited special issues for CACM, DATABASE, Journal of Global Information Management, Journal of Organizational Computing and E-Commerce, Human Resource Management, and CAIS. He served as co-program chair for the 2010 ICIS conference in St. Louis, Missouri, and is an active member in the MIS “senior scholars.”

Salvatore T. March is the David K. Wilson Professor of Management at the Owen Graduate School of Management, Vanderbilt University. He received a BS in Industrial Engineering (1972) and MS and PhD degrees in Operations Research (1975, 1978) from Cornell University. His primary research interests are in the areas of Design Science, Conceptual Modeling and Business Intelligence. He has served as the Editor-in-Chief for ACM Computing Surveys, an Area Editor for Information Systems Research and an Associate Editor for MIS Quarterly, Decision Sciences Journal, Journal of Database Management, and Information Systems Frontiers, among others. He has been involved in organizing and program committees for numerous conferences including the International Conference on Information Systems (ICIS), the International Conference on Conceptual Modeling (formally the Entity-Relationship Conference) and the Workshop on Information Technologies and Systems (WITS).