



# On the Merits and Limits of Replication and Negation for IS Research

**Sebastian Olbrich**

European Business School (EBS), Wiesbaden, Germany

*sebastian.olbrich@ebs.edu*

**Ulrich Frank**

University Duisburg-Essen, Essen, Germany

*ulrich.frank@uni-duisburg-essen.de*

**Shirley Gregor**

The Australian National University, Canberra, Australia

*shirley.gregor@anu.edu.au*

**Fred Niederman**

Saint Louis University, St. Louis, USA

*niederfa@slu.edu*

**Frantz Rowe**

Université de Nantes, Nantes, France and

SKEMA Business School, France

*Frantz.Rowe@univ-nantes.fr*

## Abstract:

A simple idea underpins the scientific process: All results should be subject to continued testing and questioning. Given the particularities of our international IS discipline, different viewpoints seem to be required to develop a picture of the merits and limits of testing and replication. Hence, the authors of this paper approach the topic from different perspectives. Following the ongoing discourse in neighbouring disciplines, we start by highlighting the significance of testing, replication and negation for scientific discourse as well as for the sponsors of research initiatives. Next, we discuss types of replication research and the challenges associated with each. In the third section, challenging questions are raised in the light of the ability of IS research for self-correction. Then, we address publication issues related to types of replications that require shifting editorial behaviors. The fifth section reflects on the possible use and interpretation of replication results in the light of contingency. As a key takeaway, the paper suggests ways to identify studies worth replicating in our field and reflects on possible roles of replication and testing for future IS research.

**Keywords:** Philosophy of Science, IS research, replication, negation, testing

The manuscript was received 10/17/2015 and was with the authors 4 months for 2 revisions.

*"Any piece of knowledge I acquire today has a value at this moment exactly proportional to my skill to deal with it. Tomorrow, when I know more, I recall that piece of knowledge and use it better."*

Mark Van Doren (1894-1972)

## 1 Introduction

An enlightened society has changed today's world in almost every aspect of daily life; overwhelmingly for the better. This change has been possible due to the achievements of modern research that is built on a basic principle that underpins the scientific process: Results should be subject to challenge and validation. That simple but powerful idea has generated a vast body of knowledge. Yet, the scientific community is increasingly faced with the allegation that science has lost its ability for self-correcting and delivering sound results as modern scientists are suspected of doing too much trusting and not enough verifying—to the detriment of the whole of science, and of humanity. "*I see a train wreck looming*", Nobel laureate Daniel Kahneman (2012) recently warned in an open letter. This development neither reflects minor cases nor is limited to a single discipline (Yong, 2012). For instance, a rule of thumb among biotechnology venture-capitalists is that half of published research cannot be replicated. That number may even be optimistic. In 2012, researchers at Amgen, a biotech firm, found they could reproduce just 6 out of 53 so called "landmark" studies in cancer research (Begley & Ellis, 2012). One year earlier, a group at Bayer, a drug company, managed to repeat just a quarter of 67 of what were called "important" papers (Prinz, Schlange, & Asadullah, 2011). In 2000-2010, roughly 80,000 patients took part in clinical trials based on research that was later retracted because of mistakes or improprieties (2014). Arguably, much of the this ongoing research is too far from the market to put people's lives at risk and some evidence suggests that the quoted crises have been overstated (Gilbert, King, Pettigrew, & Wilson, 2016). However, this situation is troubling. How much trust can we put in published results and how much need is there for replication and publication of the results of replication research, whether confirmatory or disconfirmatory, in information systems (IS)?

In the light of ongoing discussions in neighbouring disciplines and the fact that the Association for Information Systems (AIS) has released a new journal, *AIS Transactions on Replication Research* (<http://aisel.aisnet.org/trr/>), our aim is to reflect on the implications of the replication debate for the IS discipline. We are aware that a single paper is able to provide neither a comprehensive picture nor a tutorial on how to conduct replication research. Rather we present a set of individual positions regarding the value and limits of replications for IS.

The collection of the material for this paper started with a panel session at the *23rd European Conference of Information Systems (ECIS) 2015* in Münster, Germany. Rather than a mere transcription of the session, the individual standpoints of the panellists have been further developed; including reflections on remarks that arose during the panel discussion. We consider the diversity of viewpoints to be a significant contribution of this paper. The structure is as follows:

First, Sebastian Olbrich will summarize the ongoing developments, project them on IS research, and raise points of discussion. Next, Shirley Gregor will present challenging assumptions in the replication debate in relation to different types of replication in IS. Then, Fred Niederman reflects on the use and interpretation of replication results applying a broader understanding of what can be considered as replication. Then, Frantz Rowe reflects on the benefits and limitations from replication, testing and negation in IS, taking an editors and social demand viewpoint. Finally, Ulrich Frank questions whether replications in IS research can be a contribution to scientific progress in IS or rather whether they represent a reproduction of triviality; arguing that for replication to be of value one needs to build on a strong body of theory. The paper concludes with some additional remarks motivated by reflecting on the panel discussion and our thoughts moving forward.

## 2 The status of replication research in IS

The rapidly changing nature of the information systems discipline makes the development of a cumulative tradition, and the replication of findings particularly challenging: part of the so-called "anxiety discourse" that has dogged the discipline since its origins in the 1970's (Keen, 1980). Hence, IS has established a tradition that builds on the transfer and extension of existing knowledge rather than falsifying. Though the discipline has experienced impressive development, prominent members of the discipline consider greater attention to replication and verification of findings (Collins, 1985) to be essential to the race for credibility

(Weber, 1997), legitimacy of explanations (Frank, 2006; Lyytinen & King, 2004), literature reviews (Rowe, 2014) and the process of theory building. Replication and verification must involve (amongst other things) making results and underlying data more transparent. In this way results can be re-examined by other researchers and exchanged with neighbouring disciplines, thus providing for: robust verification, replication, and meta-analysis studies; opportunities for publication and debate of negative findings; and outlets for publication of high quality replication studies. “*The benefit is twofold. First, falsification increases knowledge of existing theories’ boundary conditions. Second, it also accelerates the exchange of opinions and interactions between researchers, thereby strengthening the research community. Together with inductivist research that expands the applicability of existing theories, falsificationism provides a basis and direction for new discoveries.*” (Salovaara & Merikivi, 2015).

Yet, testing and replication studies are widely considered (rightly or wrongly) to be un-publishable, with a perception that editors looking for theoretical contribution will reject replication papers because ‘this research has already been done’. Oddly, failure of replications or even negative results are difficult to publish, also; with many existing theories acting as Kuhnian “invisible colleges” where failure to replicate previous results is considered to be a problem of the researcher, not of the theory. This is not an entirely new issue as replication, verification, negation and generalisability are closely related issues. Nevertheless, valuable and recognizable examples of replication can be found within the IS field (e.g. Adams, Nelson, & Todd, 1992) as well in neighbouring disciplines (e.g. Hubbard & Armstrong, 1994). A recent, high-profile paper provided a detailed analysis of the limits of generalisability in IS research, and a process for establishing the degree of generalisability for a study (Peter Seddon & Scheepers, 2012), which seems to be the core of the problem.

Within the IS field, we tend to suffer from a degree of double-think with regard to: a) the construction of our research model and b) the verifiability and degree of generalisability of research findings (Lee & Baskerville, 2003, 2012). This leads to two observations: On the one hand, authors should be as sensitive to context as they are attentive about making modifications to instruments and theories to suit the context of their study. In the process, however, one limits the degree to which a study can test previous studies. In principle, IS papers should dutifully outline the study limitations and the context to avoid over-claims of generalisability. Last, authors may not only suggest that the findings should be verified, but should provide enough material so this can actually be done. The core issue is, that, if one must be sensitive to context, this very context might be overstretched when it comes to building models that fit substantive theory. Or as computer scientist Alex Pentland puts it “*Models which can be “tuned” in many different ways give researchers more scope to perceive a pattern where none exists. According to some estimates, three-quarters of published scientific papers in the field of machine learning are bunk because of this “overfitting.”*” (Economist, 2014, p. 22). On the other hand, once the study has been published, the limitations of the study miraculously disappear, and we “trust” the results without additional verification. Researchers cheerfully combine and extend the results of studies involving different models, samples, contexts, time periods and instruments that have never been further questioned. Conversely, results that have been almost impossible to consistently verify (let’s say for instance, five ServQual dimensions) continue to be relied on. In short, it is little understood what qualifies for a study that is worth replicating and where exactly the boundaries lie between fitting a study to context and its extension.

As is readily apparent, the broad consideration of replication studies in IS leads to an array of more specific questions and issues. We present some of these below.

One possible explanation for the relatively small number of replication studies may simply be due to the increasing competitiveness within the world of science. In the 1950s, when modern academic research took shape after its successes in the Second World War, it was still a rarefied pastime. The entire club of scientists numbered a few hundred thousand. As their ranks have multiplied, to an estimated 6 to 7 million currently active according to latest recent estimate (OECD, 2013), scientists might have lost their taste for self-policing and quality control. The obligation to “publish or perish” has come to rule over academic life (Gad-El-Hak, 2004). Therefore, the competition for jobs, particularly among young researchers, is getting increasingly intense. Particularly for young, untenured researchers, it seems to be a viable strategy to mainstream at the cost that results become rather trivial, not to say boring.

- Replication does little to enhance the career. In order to safeguard their exclusivity, the leading journals impose tough reviews and high rejection rates – of more than 90% of submitted manuscripts in top journals. Little wonder that one in three researchers knows of a colleague who has pepmed up a paper by, say, excluding inconvenient data from results “based on a gut feeling” (Fanelli, 2009). What is observable often implies a “hand washing in statistical

procedures" (McCloskey, 1985, p. 18); even if no wrongdoing is imputed. Possible fields of tension might be the value of replication and negation in terms of journal publications (Kane, 1984) and possible incentives for (junior) researchers to get engaged without compromising their integrity towards senior scholars (Dennis & Valacich, 2014).

- The benefits of learning from failure have been recognized on various occasions (e.g. Arin & de la Torre, 1998; Edmondson, 2011; KC, Staats, & Gino, 2013). Having a background in natural sciences myself (Olbrich), I consider the ongoing questioning of existing theory and repetition of already conducted experiments to be a cornerstone of scientific understanding (Popper, 1980). To me the discussion about and interpretation of why experiments succeeded or failed is one of the most productive and valuable means of learning. That being said, in the pursuit of new theoretical understanding at the intersection of rapidly developing technology and the context in which organizations operate, the Popperian understanding of the scientific process is challenged and complemented by inductive approaches. Therefore, I do not see why such inductive studies should not be subject to testing and verification in order to learn for future. Given the strong ties of IS and engineering, the discussion about failure to replicate seems equally valuable; particular when investigating cases.
- Failures to prove a hypothesis are rarely even offered for academic publication, let alone accepted. "Negative results" now account for only 14% on average (numbers depend on the discipline) of published papers, down from 30% in 1990 (Fanelli, 2012). For the scientific discourse, knowing what is false is as important to science as knowing what is true. The avoidance of reporting failures means that researchers waste resources exploring blind alleys already investigated by other scientists instead of for new developments. Yet, incentives to do so are rare.
- For practice, reporting on failed experiments seems to be equally valuable as organizations face the chameleonic nature of context on a daily basis. Naturally, the issue arises that few institutions (public or private) are interested to share data in failed experiments with technology; leaving researchers with only limited access to failed cases. Yet, it might also be that the theoretical core of the IS discipline is simply not strong enough to hold up for rigorous testing. Hence, fundraisers and project sponsors might simply see too little value in pursuing that line of research.

### 3 Types of Replication Research and Associated Challenges

The field of IS is recognized as having both technical and social aspects, concerning both information technology and the behaviour of individuals and communities as they interact with technology. Many researchers in IS have varied backgrounds that mirror this range of phenomena, including social, natural and technical sciences. My own background (Gregor) includes an undergraduate science degree with majors in mathematics and psychology. This early training was no doubt the initial basis for an ongoing belief that repeated questioning and testing of knowledge claims is a vital component of scientific endeavour (although not the only component). However, this belief does not mean that some assumptions behind the current debate on replication should not be challenged.

In studying psychology, it was common for classes to participate in informal experiments so that they could observe replication of some well-known effects, such as the Stroop Effect (Stroop, 1935) and the limits of short-term memory. Anyone who has studied psychology will also recall the almost obsessive attention to detail required in recording the experimental method so that other researchers could repeat and verify the results reported (see APA, 2008). My subsequent studies included information technology and systems and then employment in software engineering. Here again there was an expectation that it was desirable for knowledge to be as reliable as possible, as the basis for action in designing and developing systems. Designs of systems, processes and algorithms need testing and investigation of use that is as thorough as possible to lead to reliable actionable knowledge. But is this desire to see knowledge claims challenged and tested the same as asking for "replication research" in IS?

The situation with much research in IS that purports to be “scientific” is perplexing (Hempel, 1966; Toulmin, 1958)<sup>1</sup>. Some practices are in contrast to what the naïve undergraduate learns as the canons of science: for example, research models are revised and hypotheses removed after the results are obtained; theoretical underpinnings for a research model are changed at the end of a study; details of empirical work such as the date and country in which a sample of participants was obtained are omitted; and the boundaries within which a research model is expected to hold are not stated. In addition, some methods are bedevilled by well-known methodological issues (see Gregor & Klein, 2014). Replication of studies would appear desirable, but the worth of replication under these conditions is debatable.

The manner in which replication is called for in IS should be carefully examined, especially given the distinguishing characteristics of the field. IS is “harder” in that it deals with complex IT-based artefacts when compared with fields such as management, and yet it is “softer” in that it has to deal with the complexities and uncertainties of human behaviour compared with the physical sciences. IS thinkers are prominent in promoting Design Science Research and critical thinking about the philosophy of technology. It may be that IS, belonging wholly neither to the natural or the social sciences, can add to the debate on replication research in ways that other fields cannot.

From this point of view, there are several issues that I believe pertain to replication research in general, that should be kept in mind:

- Is exact replication possible? Given the open nature of information systems many of our studies are highly context dependent and exact replication of the context will not be possible. This issue is discussed further below.
- How much is enough? How many repeated confirmations do we want? Should we only publish work that disconfirms some prior finding? Given the probabilistic nature of much IS research, is it even possible to disconfirm prior findings? This issue has been extensively discussed in the philosophy of science, including by Popper (1980).
- Is it worth publishing replications of work that made claims that were not very interesting in the first place, perhaps being almost tautological, as in the statement that an artefact that is perceived as being useful is more likely to be used? Ulrich Frank makes a similar observation below.
- Are replications worthwhile if they have methodological biases similar to those in an original study? For example, some cross-sectional surveys are recognized as problematic because of common-method bias and so on. What is the value in a further study that has the same problems? Colquitt and Zapata-Phelan (2007, p. 1284) distinguish between “operational replication” and “constructive replication”, where the later creates a more stringent test by avoiding limitations of the original study’s methods.

My further discussion is framed around the three types of replication research identified by the editors of TRR in its initial issue (Dennis & Valacich, 2014)<sup>2</sup>. There are nuances relating to the place of replication research depending on the type. Paying attention to these nuances and the respective roles of theory testing and theory building may assist authors, by leading to a broader discussion and understanding of what replication research can and cannot accomplish.

### 1) Exact Replication

Exact replication is said to include research articles that “are exact copies of the original article in terms of method and context. All measures, treatments, statistical analyses, and so on are identical to those of the original study” (Dennis & Valacich, 2014). It may be helpful to tease out what is required here a little more. The editors no doubt did not mean that truly “exact” replications are expected. As pointed out by Tsang and Kwan (1999), whom the editors cite:

*Even in the natural sciences it is impossible to impose exactly identical conditions when repeating an experiment. For example, the colour of an instrument, the time of performing the experiment ... and so on are rarely identical. ... a more modest concept of replication is called for. ..., if most social*

<sup>1</sup> Note that the philosophical position underpinning my part of the discussion is one of a non-naïve realist, a non-relativist and a non-positivist. It is believed that some analysis, arguments and theory are to be preferred to others, because they can be better justified by reference to supporting evidence and relevant existing knowledge.

<sup>2</sup> The debate on replication research in the social sciences notes a confusing array of terminology for replication research (see Tsang and Kwan 1999). However, the definition by Colquitt and Zapata-Phelan (2005, p. 1284) is pertinent: replications “offer neither new concepts nor original relationships”.

*conditions that are likely to be relevant are controlled for, we can say that a replication has been performed in the social sciences (p. 763).*

However, given that exact replications are not feasible, repeated tests of proposed knowledge claims (theory testing) are important in developing reliable knowledge. Results that were due to some extraneous factor such as peculiarities of the sample selected should be detected and discounted. Many writers in the philosophy of science recommend repeated tests of knowledge claims in this way (e.g. notably Popper 1980). When a hypothesis or knowledge claim is repeatedly confirmed in approximate replications, then using inductive reasoning we feel that our knowledge has a sounder basis (see Hempel (1966), but not Popper (1980)). If TTR can prevail against current prejudices against replication work then IS knowledge could be given firmer and more reliable foundations.

One issue, however, still needs to be recognized with the “almost exact” type of replication. How closely can we approximate an “exact” replication in IS? So much work is context-dependent that it is difficult to think of effects that we would expect to hold across repeated tests. Perhaps some of the principles in human-computer interaction qualify, such as the design of screen features to match human short-term memory limitations as may some of our more technical knowledge such as modelling methods. Work in more complex social settings may not be even approximately replicable.

## 2) Methodological Replications

*“These articles use exactly the same methods as the original study (i.e. measures, treatments, statistics etc.” but are conducted in a different context” (Dennis & Valacich, 2014).*

Work of this nature may be better thought of as offering opportunities for theory refinement. That is, a failure to replicate in a new context can suggest ways in which theory can be modified to take account of different factors in the new environment (as noted by Dennis & Valacich, 2014). In the language of Colquitt and Zapata-Phelan (2007), this type of work can lead to subsequent studies with a moderate level of further theory building, where articles introduce a “new substantive mediator or moderator of an existing relationship” (p. 1284).

This work can also be highly significant for design science work. Understanding more precisely how artefacts are used in different contexts allows us to extend design knowledge. Unfortunately few examples spring to mind of publications of this type in IS. An example from medical science illustrates how important attempted replication of artefact usage can be.

*“In the late 1990s, a team of researchers in Boston headed by Dr Judah Folkman reported that a new drug called Endostatin shrank tumours by limiting their blood supply (Folkman 1996). Other respected researchers could not replicate the effect even when using drugs shipped to them from Folkman’s lab. Scientists eventually replicated the results after they had travelled to Folkman’s lab to learn how to properly manufacture, transport, store, and handle the drug and how to inject it in the right location at the right depth and angle.” (Shadish, Cook, & Campbell., 2002)*

The suggestion is that TTR could add further value to our discipline by welcoming replications of behavioural and design science work in new contexts, with the aim of theory refinement and development.

## 3) Conceptual Replications

*“These articles test exactly the same research questions or hypotheses, but use different measures, treatments, analyses and/or context” and are the strongest form of replications (Dennis & Valacich, 2014). The importance of work of this type is supported by philosophers of science. (Hempel, 1966) says “the confirmation of a hypothesis depends not only on the quantity of the favorable evidence available, but also on its variety: the greater the variety, the stronger the resulting support”.*

Studies published in TTR of this type include Asamoah, Andoh-Baidoo and Agyei-Owusu (2015) who replicated a study of ERP implantation in the USA in Ghana, and found that the original results did not hold, presumably because of the different context. The authors claimed that their study was a conceptual replication, but it may be better thought of as a methodological replication.

Following in the spirit of Hempel, TTR could be bolder and publish studies that are replications in the sense that they test the underlying argument of some well-known theories, but use entirely different methods from those used previously. Theory that is founded on many survey studies could be tested in experiments, or be matched against secondary data from public databases or large commissioned

empirical studies. For example, Nansen et al (2012) report a large study of the adoption rate of high-speed broadband. The factors that are shown to be influential in determining adoption (household composition, ownership, and income) could be compared with those expected from theories of technology adoption in IS. Questioning of conventional wisdom in this way may not be acceptable in mainstream outlets but could lead to better grounding of IS theory.

In sum, my contribution to the debate on replication studies in IS stresses the following. Replication is important in principle, but even approximations to exact replication may be difficult in IS. Re-testing of findings in different contexts is important for the opportunity for refining theory. There is still a question of whether one more confirmation of well accepted findings adds further value and that replication that looks for a diversity of evidence is preferred. Finally the field should be especially open to efforts to test, re-test, and extend design science work.

## 4 A Broader Understanding of Replication Research

The term replication can be used strictly to suggest the repetition of a study with zero or near zero variation from the original. Alternatively it can be used to indicate a study that originates with a completed published study and repeats that with some amount of identified and explained divergence – some of which may be intentional. Following (Berthon, Pitt, Ewing, & Carr, 2002), such a broadly conceived replication study can vary on four dimensions which are: problem, theory, method, and context. Such replication can vary in the amount of divergence from the original on any of these dimensions as well as on the number of dimensions where there is divergence. As the amount of divergence from the original grows, the study approaches becoming a “new” study rather than an example of replication. Unfortunately, a sharp clear line where such transition from replication to not-replication has not been established, so we are left with a continuum where at each extreme we have clear replications and clear new studies, but in between we have a “grey area” open to interpretation and preference. Perhaps at some point a clarifying definition or sufficient examples will emerge to make the dividing line clearer.

My own view (Niederman) leans toward defining replication broadly. As I see it, all studies from the “pure” replication to the point where there is no longer any reasonable justification for a replication study should be defined as replications. Thus my views on the importance and role of replications are based on this broad view. This view is supplemented with the understanding that the accumulation of knowledge relative to a particular IS domain or problem is more important than the classification of any particular study in terms of whether or not it is a replication. The overall issue pertains to how the IS field can effectively build a reliable and useful body of knowledge. I see the tactic of replication of prior studies as one among many approaches to such building of IS knowledge.

From this perspective, I consider replication studies to have significant potential for benefitting our MIS research domain. To the extent that anyone could conceive and execute a “pure” replication, as one might in the natural sciences, I applaud that effort and suggest that potential benefits in terms of providing a platform for evaluating the initiating study are substantial. In such an instance, should findings be identical in direction and strength (assuming a statistically based decision rule), the findings of the initial study have rightly gathered additional support and are worthy of more confidence if there is an opportunity to apply them. If the findings are not identical, new questions can be raised about the nature of the phenomenon, the inclusiveness of influential factors, and the detailed execution of the study. Where findings vary modestly from the original study, perhaps some contingency or fine tuning will be required. Where the findings are widely disparate between studies, serious questions about the underlying phenomenon, theory, and research approach will need to be raised. Note that at this point if the domain is indeed significant such questions, while troubling in the short run, offer the prospect of guiding scholars to better and more robust eventual answers. If the domain is of modest significance, the topic can be dropped in the short run but remain for revived consideration should circumstances warrant in the future. Whether or not such a perfect replication is even conceivable in MIS or any behavioural or “artificial” science is outside the scope of this discussion.

As the strictness of study components loosens, such that problem, theory, method, or context varies from the original, the interpretation of results becomes more complex. In principle four major alternatives can be seen: (1) the initial study was correct and the follow up misguided; (2) the follow up is correct and the initial misguided; (3) both are misguided and perhaps in different ways; and (4) both are correct but likely in different ways. For this last alternative a new construction of knowledge could be considered. A blueprint showing plumbing may be substantially different from one showing electrical conduits yet both

may be correct (or correct enough) to lead to successful building. It would be in the interest of the scholars presenting such findings to propose one or more explanations for the contrasting results. Over time a sequence of replications may show a predominance of one or the other findings, may show additional alternatives, or may reveal differences in timing, setting, population/sample, or other factors that lead to more nuanced understanding of the phenomenon under study including the nature of constructs and their relationships.

Such interpretation complexity is shown in the recent discussion of a set of studies to test published psychological findings (Carey, 2015). Among 100 replicated studies most frequently observed findings were those pointing in the same direction but without as much strength as the published results suggested. No definitive answer is proposed as to why this array of findings was observed. There may be one universal explanation: for example, that without conscious intention scholars tend to stack intangible procedural decisions in favour of stronger results. Alternatively, underlying theory may be incomplete or samples may vary in other than normal distributions. What is most interesting based on these replication experiments is how they generate the opportunity to re-evaluate the meaning of the initial studies and to update, sometimes strengthen, and sometimes reject or modify the original conclusions. Of the 100 initial findings reinvestigated, perhaps only a subset have important enough implications to be worthy of further study, but over time we might expect all of these to initiate new investigations with the prospect of generating deeper understanding of the original domain of interest.

The great opportunity offered by replication is its potential to assist in the creation of a collective body of knowledge based on an accumulation of evidence. We have seen this to some extent in the integration of prior literature embedded in most published studies. Authors will review prior findings on a topic, but the emphasis on new discovery may push them away from simply rechecking the initial results and toward assuming that initial studies are correct and robust in an effort to extend them. A pitfall of this latter approach is the potential of creating a long chain of logical formulations that rest on disputable foundations.

A theory that has been tested extensively, using only one method, say experiments or surveys, should stand up to observation in case studies and grounded theory – or vice versa. Should, for example, usefulness lead to more use, as observed in particular individual cases and commentary by experts, the confidence and applicability of the theory would increase strongly. Note also that a correspondence between a variance theory showing correlation of initial states with outcome states and a substantive theory showing the particular dynamics or processes by which such initial states move to particular outcomes can form a much more powerful understanding of the phenomenon than without this correspondence. This situation would be analogous to an initial variance level understanding that when we take aspirin we tend to lower fevers, which is useful knowledge for health care providers. Supplementing this knowledge with the actual chemistry by which the components of aspirin combine with physiological conditions can allow for much more targeted application, understanding of exceptions, and consideration of new and more effective alternatives based on underlying principles. In IS if there is a reliable variance relationship between usefulness and use, understanding the dynamics by which users see and exploit that usefulness might help practitioners in situations where substantial numbers of users do not initially see the usefulness, the experience of usefulness is not obvious from actual use, or where particular impediments to applying the technology to tasks need to be resolved.

If alternative approaches do not yield consistent findings, the results should lead to deeper questions. Are there contingencies omitted from initial models – can looking at results from replications lead to a more thorough taxonomy of entities that are influential in creating particular outcomes? By examining different findings over time, can we infer underlying trends or trajectories that explain or help us predict future states? In the short run, the unearthing of countervailing observations may lead to a state where it appears that the field has no theory at all as conceptualizations appear to be refuted. It is an open question as to whether persistent observation will lead to universal or deterministic theory, but I am confident it should lead to increasingly strong domain knowledge with the prospect of a theory base that is both reliable and useful.

## 5 Issues in Publishing Replication Research

Following the reflections in prior sections, I would argue that we first need to discuss “what is replication?” Spontaneously we often think that it means verifying that the empirical findings are the same (versus the theory). Espousing this empirical view as my starting point, the narrowest definition I can consider for

replication is the ability to address a problem by reproducing or verifying the same (model, theory or principles) with different data originating from a different sample from the same population in the same spatio-temporal context. By analogy to Tsang and Williams' (2012) typology of generalization, this type of replication would correspond to "*within population replication*". Needless to say, if the problem is significantly different there is no point in referring to the work as replication. The narrow within-population replication is well suited when the population has a large number of individuals so that several draws (samples) can be constituted with little bias. The case of little bias could apply if we were interested in the US population of individuals as IT adopters, but would be much more difficult, if not impossible, if we consider the population of CIOs of large firms in France. Despite an apparently narrow scope, if a problem is important for some stakeholders or ultimately society or a nation, such studies are needed to avoid misconceptions and to adjust plans and policies. Keeping the same context, "*cross population replication*" (id) would typically examine results for a different population (e.g. generation X vs Y), bringing up issues about why and how results might be different. Changing context, comparative studies researchers have to make assumptions about the similarities and differences across contexts as we often do in international studies (Warkentin, Charles-Pauvers, & Chau, 2015): for instance by drawing on Hofstede (Preston, Karahanna, & Rowe, 2006) or other international differences such as climate and wealth (Chen, Po-An Hsieh, Van de Vliert, & Huang, 2015) to derive an explanation for other differences and similarities. This type of replication is "*across context replication*" (ibidem), be it spatial, as in our example, or temporal.

Beyond types of replication distinguished by analogy to Tsang and Williams' (2012) generalization typology, we also identify "*contrarian methods replication*". Labeled after the series published in the *European Journal of Information Systems*, contrarian papers such as Hunter (2010) use the same data (context and sample) and theory for the same problem but different methods. This work can successfully correct errors and challenge some practical and theoretical findings across contexts in previous findings (Koufaris, 2010).

By extension of the problem space we might also consider another type of replication, "*solution replication*", which has potential in design science research. A solution developed in a design science project should be aimed at a class of problems—not just at one specific problem. If this class of problems cannot be defined formally, it is not possible to prove that it will work for all instances of the class. Therefore, it can make sense to test the solution by applying it to selected instances of the targeted problem class. A further test of the solution may be applied to a further problem of the same class—which could then result in a refutation of the claim that the solution works for the entire class. For example, a solution might be a tool for business process analysis. A first test could target problems of business process analysis in company A and report the successful application of the tool. A further test could focus on other problems of business process analysis in company B, within the same class of problems. However many of us would still be reluctant to call this replication, because, as I see it (Rowe), our notion of replication depends on the existence of a theory, although others may debate this view. Often, cases like the one above lack a convincing theory. Therefore, every simple test results in a snapshot of a moving target only. Hence, a further snapshot will represent a different subject (lack of commensurability).

With such considerations we see that all types of replications question epistemological conditions for knowledge reuse respectively in the same population, in a different population, a different context, with a different method and even possibly within a class of problems. All the above types of replication are all the more important when the original study has had high impact; because retesting the findings under different conditions contributes to refining the conditions for the validity of new knowledge.

However, contrary to my starting point, we can also consider replication as the ability to reproduce the theory and to generalize it, à la Yin, rather than as the ability to replicate the empirical findings. In science we research and advance knowledge by iterating between the phenomenon and the theory. This view has been firmly justified since Kant. However, we give primacy at some point (typically at the starting point) of the research process to one or the other, and can even push to the edges (Grover & Lyytinen, 2015) with genres that focus on one or the other (Rowe, 2011). Typically empiricists would start from the phenomenon arguing there is no possible knowledge claim without experience. But rationalists would take the opposite stance (Leibniz, 1765). Regardless of our inclinations, this means that we learn by observing about what happens and by reflecting about possible futures with our thinking and imagination. For our object here, it is thus both important to observe practices and interpret meanings in different contexts so that we make new inferences about the validity of our claims when the population and/or context change, thus leading to either different solutions to the problem, if we think the theory still holds, or to a revision or rejection of the theory. In fact, theory could predict different findings in new conditions. In a position à la

Yin, we consider that replications can be useful if they allow refinement of boundary conditions for the theory (Seddon & Scheepers, 2012; 2015), *whether findings are replicated or not*. If we focus on replication of empirical findings, as noticed by Ulrich Frank in this paper, simple confirmation or verification might not be considered interesting enough for publication. As an Editor I would disagree with such an attitude, if a clear explanation is provided for why it is theoretically surprising that similar research findings are found. Unfortunately this is what is too often lacking, and motivates rejection; that is, a lack of theoretical discussion of confirmatory findings.

If our observations – typically in case studies – and our thinking ability enable us to generate and reproduce new potential behaviours, we can also simulate futures. This is an additional reason for viewing the interest of replication, both real through comparative observations and artificial through simulation. In this sense, we can embed learning in simulation programs and allow both for replication and change, but in a broader sense à la Lee and Baskerville (2003). We could then refer to “*embedded replication in simulation*”. Regardless of these subtle nuances and terminology issues, I deeply regret that in Information Systems we publish so few simulation studies (Spagnoletti, Za, & Winter, 2013). I think this is also related to the difficulty of opening the black box to the mundane IS reader when publishing such studies, since paradoxically, there is no difficulty in replicating them if the code is made available to other researchers (Collberg & Proebsting, 2016). The point would be to develop theoretical arguments for calibration changes when doing across-contexts replications.

Replication is also fundamental for literature reviews, especially meta analyses and I join Seddon and Scheepers (2015) in disapprobation of Tsang and Williams (2012) when they decided to exclude meta-analyses from the scope of their framework for generalizing research knowledge claims. In fact the power of meta-analysis is to systematically compare results on a problem and thus discuss the power of findings across cumulative evidence. Such endeavours are extremely important for convincing and legitimizing stakeholders when intervention on a phenomenon is considered such as in medicine or when we design an artefact for solving a particular problem. More generally we can only draw solid conclusions from literature reviews about the studies they synthesize if there is at least a minimum level of replication across studies that are considered and if a high degree of systematicity in the review method is sought by the researchers performing the review (Rowe, 2014). This is all the more important if these literature reviews are theory testing reviews (also known as meta-analysis). In this case we can also recommend 1) that publication bias be limited – e.g. by avoiding selecting only the basket of eight journals – and 2) that papers with non-statistically significant results (i.e. non rejecting the null hypothesis) be searched for and included in the review. This last task is difficult because it is well known that top journals generally avoid publishing negative results (Eisend & Tarrahi, 2014). Therefore the launching of the *AIS Transactions on Replication Research* is particularly suited to advance certain types of research and genres such as theory testing reviews.

In particular, it is surprising to note that negation is undervalued. Negation is the practice of rejecting a hypothesis. During my involvement at the *European Journal of Information Systems* as Editor or Editor-in-Chief (5 years) and at *Systèmes d'Information et Management* as Editor-in-Chief (13 years), I noticed that papers were rejected if too many hypotheses were rejected. In a Popperian epistemology where science advances by rejecting hypotheses, this may look as a paradoxical failure of disconfirmation. However this is not so paradoxical if one accepts that to advance science such hypotheses need to be bold and interesting and should hold until rejected in a new context or with new boundary conditions. Still too often reviewers recommend rejecting these negation papers on the basis that (too many) hypotheses are deemed unsupported, and not because they are not interesting. Rejecting an interesting (either novel and well argued, or deemed valid) hypothesis plays in fact a crucial role in a positivist epistemology.

I have already mentioned a number of difficulties or limitations that would make replication work less interesting and publishable. I would like to end my short comments by articulating also why I find valuable the view of those who, like Ulrich or Shirley, think that replication may not really be needed, or only in very specific circumstances. Most replications are testing the confirmation of a kind of regular association à la Hume. This (dis)-confirmation is itself highly dependent on the possibility of viewing the phenomenon as a closed system (Sayer, 1992), a possibility that is rarely observed in social sciences and in Information Systems, as noted in this paper by Shirley Gregor. In social worlds and open systems such as those encountered daily life, changes are permanently attacking order and replications (may) become(s) impossible to achieve without considering hidden forces and generative mechanisms that sustain order, an epistemological position which is leaning towards interpretivism and/or critical realism. In other words the replication and negation discourse might be limited to those who espouse a positivist view of science

(Lee & Baskerville, 2012). In addition, accumulation of research on generative mechanisms by which social relationships are maintained is useful for a better understanding and explanation of the social world (Avgerou, 2013) and need not depend on an ability to calculate their effects quantitatively (Sayer, 1992).

## 6 A Sceptical View on Replication

When I received the invitation to participate in this debate my first reaction was to welcome the demand for replication studies. I thought it might be a worthwhile approach to effectively challenge some misconceptions of research in IS. The following considerations are mainly based on a narrow conception of replication that is related to neo-positivist research methods, since it seems to me that the current debate in IS was inspired by discussions in the natural sciences, in medical science and psychology, that is, disciplines where neo-positivist methods are clearly dominant (and often impressively successful). Therefore, I agree with Frantz Rowe's speculation that "the replication and negation discourse might be limited to those who espouse a positivist view of science."

The "theory fetish" (Avison & Malaurant, 2014) in IS was promoted by a research model that to a large extent is built on simplified ideas of logical positivism and of Popper's conception of scientific progress through refutation. First, it assumes that there are theories suited to explaining relevant phenomena. That assumption implies the applicability of the Hempel/Oppenheim schema which requires a clear distinction of *explanans* and *explanandum* or, in other words, a concept of intended causality. Second, it assumes that *explanans* and *explanandum* can be operationalized and mapped to corresponding independent and dependent variables, which in turn can be used to collect the respective data. The first assumption is problematic for two reasons. First, there is no precise conception of theory that would permit researchers to clearly distinguish theories from other linguistic constructions that are not theories, which is remarkable for a discipline that regards the quest for "rigor" as one of its cornerstones. The actual use of the term *theory* in our field not only gives the impression that there are "conflicting notions of theory and theoretical contribution" (Markus, 2014, p. 342), but also that theories are the result of a social construction rather than an expression of an epistemological quality. The second assumption is challenged by the concern that formal models of complex social phenomena will usually include simplifications that compromise the validity of the respective studies, and that these doubts about validity cannot be overcome within the positivist paradigm alone. Both assumptions have been the subject of a long ongoing debate in the sociology and philosophy fields. They are also reflected in the debate on quantitative versus qualitative methods in IS (which is based on a misleading use of terms, since quality and quantity are orthogonal dimensions). At the same time, there are apparently few theories that fit the positivist idea of explanation and prediction. I assume that there is a valid reason for this: The subject we are dealing with—action systems that interact with information systems—is characterized by *contingency*. It does not matter whether this assumption is grounded ontologically (a property of the social world as such) or epistemologically (caused by restrictions of human recognition). Obviously, there is no way to prove such assumption. However, there is little evidence so far to invalidate it either. It is mainly the second aspect that could be directly targeted by replication studies. If a construction termed theory is aimed at a contingent phenomenon, a particular model of the theory that is used to test it will result in only a snapshot of a moving target. Therefore, a replication study is likely to produce a different snapshot, which would not only contribute to the refutation of the theory within a neo-positivist world view, but might also foster a critical discussion of inappropriate conceptions of theories in general. This is the outcome I was thinking of when I initially embraced the idea of replication studies.

However, when I thought again, I had to revise my first impression. A review of empirical studies published in leading journals would unearth evidence for an even more pessimist suspicion. A remarkable number of studies—I would dare to say the majority—lack a convincing theoretical foundation. Instead, they are based on constructions that almost equate to tautologies, which is in fact a reflection of the contingent nature of our subject: The only way to avoid immediate refutation is to reduce the information content of the core propositions. I do not blame the authors of the respective studies. It seems they are the result of the neo-positivist research program that at best came with the good intention to replace "a mishmash of fuzzy thinking and incomprehensible jargon" (Dearden, 1972, p. 90) with hard evidence, and that was established as an outcome of the "race for credibility" (Weber, 1997, p. 2). Unfortunately, research programs that oversimplify the IS environment lose the ability to account for peculiarities of social systems, particularly their emphasis on contingencies of action, systems that are complex, and the idea of voluntarism, purpose, and intent. Excluding these characteristics from IS theory may be of greater consequence than excluding them in the natural science; and by the same token, relying on natural

science for models of theory and research, while understandable, may be a source of these limitations. A further argument supports this skeptical assessment of the positivist notion of theories. Information systems are linguistic constructions and the primary way to use them is through language: The language we speak has an impact on the systems we design, which will in turn influence the language we use and, as a consequence, the way we act and live. However, the only way to develop theories about the use of language is through the language we speak. It is our only instrument, but it also limits our view, because we cannot go beyond it (Maturana, 1987 p. 90f; Wittgenstein, 1961 § 5.6). One does not have to be a follower of Popper's one dimensional idea of scientific progress to realize that empirical studies without informative theories are hardly suited to producing surprising new insights, that is, studies bold enough to invite challenge. For the respective studies, replication does not make much sense, for it is likely to result in the reproduction of trivia rather than in the confirmation/refutation of serious theoretical contributions.

Related to my previous considerations is a further principal concern over replication studies in IS. The digital transformation is a fascinating phenomenon that lies at the core of our research subject. To benefit from the opportunities that transformation creates— instead of suffering its adverse consequences—it is certainly not sufficient to study present or past action systems and related information systems, and to replicate the respective studies. Instead, I would suggest emphasizing an essential aspect of the original, rationalist notion of theory ("teoría"): to transcend the factual world by contemplation. That means looking beyond current patterns of developing and using IT or , in other words, developing justified (!) models of *possible future worlds* (Frank, 2006; Rorty, 1999) that serve those who live the future as an inspiration, and a meaningful orientation. The respective constructions cannot be invalidated by confronting them with empirical data, because they are purposefully different from *reality*. They could, however, be tested; for example, through critical discourses, (thought) experiments, simulation models and suchlike. Apparently we should not be too excited about the prospects of replication studies.

Nevertheless, the chances are that we will see a growing number of such studies, if not a new wave of hype. First, they create new publication opportunities, which are sought after by many individual researchers. Second, they are suited to promote the image of respectable and responsible research. In other words, replication studies could enhance the legitimization of the entire discipline. In my view, those promises made by replication studies are suited to convince only the naïve or cynics.

It will come as no surprise that I am skeptical about replication studies, but that does not mean that I would reject them altogether. If an empirical study is based on an informative theory that has not yet been sufficiently confirmed, the different forms of replication studies described by Fred Niederman can contribute to analyzing and validating the theory. Furthermore, a research governance that includes the demand for replication studies could foster honest and responsible handling of data. I do not know of any manipulation in our discipline: but given the pressure that some researchers work under, I would not be surprised to hear of instances. The threat of being unmasked as a cheat would probably create a significant barrier for any researcher tempted to manipulate data. However, with respect to the contingency of data, different results are certainly not a sufficient indication of manipulation. Therefore, a smart cynic may not be impressed by the prospect of replication studies that produce different data.

Finally, one could also use a less constrained notion of replication, as suggested by Fred Niederman and Frantz Rowe. In that case, replication could contribute to narrow or widen the scope of a theory by the systematic variation of boundary conditions. In an even more abstract sense, as outlined by Frantz Rowe, one could regard replication as an integral part of an ongoing critical discourse. In that sense, replication would mean starting a discourse on a previous study, story, scenario, artefact, method etc. and then trying to validate it by *reconstructing* its epistemological outcome. However, I would not advocate saving the notion of replication by broadening it. Instead, it should be used only in those, I am afraid, very rare cases where there are good reasons to doubt the outcome of studies that were intended to test informative theories, that is, theories capable of surprising the experts. But as long as such theories remain scarce, I agree with Shirley Gregor (2014) that we need more theorizing in IS instead of an ongoing repetition of testing a few "kernel" theories of little epistemological value.

## 7 Concluding remarks and key takeaways

In the course of this paper we have reflected on the value of replication, verifying, testing and negation for IS research from different viewpoints. It seems that IS research might follow similar developments in neighbouring disciplines in neglecting testing and replication; i.e. doing too much trusting at the cost of credibility. We jointly conclude by supporting more testing of theory and verifying of published research

results, both of which, in our opinion, advance science and practice. However, the claim is not supported unconditionally as all authors raised some level of concern about the type of replication studies that are worth encouraging. For the IS field the question remains of what actually is replicable and what is worth replicating.

As IS cases are often grounded in social theory, the elephant in the room is combinatorial explosion. To date, no clear picture exists of how many theories might be needed to explain all IS phenomena. The current conventions of the IS field incentivise the expenditure of time on existing reference theory, rather than pushing for the development of new theory or subsets of theory. The natural price for continuing this convention is twofold: First, only a subset of 'key' phenomena is studied while others are not attended to at all, or only when they seem fashionable. Second, barriers to introducing new theoretical bodies are getting increasingly high; this holds true for indigenous theory and the application of a subset of theory. As a way forward, the role of inductive and deductive theory for IS research should be further evaluated. On top of that, it should be clarified how results from both approaches can be combined and put to the test.

IS remains a vibrant field to operate in, as it investigates the intersection of developing technologies and the increasing role such technologies play for the individual, the organization and society. A key driver for many of us is taking a leading part in facilitating the corresponding changes. Taking the role of an active designer, we constantly challenge the status-quo of the existing world and aim to transform it. Therefore, it does not seem very fruitful to test artefacts which are supposed to create future worlds against the experiences of the existing reality as we experience it. That being said, innovation itself is something that, per definition, is difficult to test. Rather than taking "snapshots of a moving target", we might find value in comparing multiple snapshots over time. Such long term studies and the repetition of experiments with maturing technology might lead to a viable knowledge base. We may need to undertake the process of developing approaches, attitudes, incentive systems, and tools for such a transformation.

It is unlikely that exact replication of the context in which an artefact is embedded is possible. Further, exaggerating the incentives to replicate studies might even lead to a replication or confirmation bias; or following Ulrich Frank's logic, it may serve as a counter to strong incentives to fudge original studies. As claims of correlation are regularly established through surveys, the question is raised as to whether an exact replication would provide any value at all. Given the common method variance issue, it may be questionable if any improvements in the correlation coefficients can be achieved or have much meaning. Following this line of argumentation results in a few promising directions: First, the generalisability of results and the strength of theory. By varying the boundary of a theory, it shifts in how generally applicable the statements are or at what strength the theory applies to specific cases. The same logic applies for replication, of course. Second, one might get the idea of conducting replications that use the same theory, but different methods and approaches. It might be interesting to see if, for instance, survey results are still valid when using qualitative methods, or whether patterns can be identified in comparing the results of studies with different approaches to testing theory.

As a way forward, a nuanced understanding of replication is advised. As it seems unlikely and questionable to exactly replicate previous studies, conducting comparable studies that investigate different contexts are worth supporting. When we begin to appreciate and investigate possible sources of difference (e.g. by publishing publications that failed to prove hypothesis), existing results can be put to the test and eventually, given a higher degree of confidence, connected to a cumulative body of knowledge. Overall more research on failure (projects, prototypes, experiments ...) and underlying causes should be encouraged. Last, it should be highlighted that it is not the number of replications that count in making a theory stronger, but the scientific contribution of the corresponding studies (e.g. in uncovering detailed justification or mechanisms for the relationships posited, or in stronger tests of these relationships). It seems advisable to focus on previous results that are worth replicating; i.e. results that are important and have impact in the world outside academia. Consequently, such a statement requires researchers to articulate theories and hypothesis in a (positivist) way that they actually can be tested and replicated.

## Acknowledgments

To the organizers of ECIS 2015 who have been great hosts in Münster, the audience of our panel discussion who engaged in a vivid debate and the organizers of the ECIS track “Advancing Theories and Theorizing in IS Research” who allowed the discussion to continue. Also, Dirk Hovorka who was one of the original members of the group and unfortunately could not make it to Münster.

## References

- Adams, D. A., Nelson, R. R., & Todd, P. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly*, 16(2), 227–247.
- APA. (2008). Reporting standards for research in psychology Retrieved 12 November 2013, from [www.apa.org/pubs/authors/jars.pdf](http://www.apa.org/pubs/authors/jars.pdf).
- Arin, A., & de la Torre, J. (1998). Learning from failure: Towards an evolutionary model of collaborative ventures. *Organization Science*, 9(3), 306-325.
- Asamoah, Andoh-Baidoo, & Agyei-Owusu. (2015). Impact of ERP implementation on business process outcomes: A replication of a United States study in a Sub-Saharan African nation. *AIS Transactions on Replication Research*, 1(4), 1-19.
- Avgerou, C. (2013). Social mechanisms for causal explanation in social theory based IS research. *Journal of the Association for Information Systems*, 14(8), 399-419.
- Avison, D., & Malaurant, J. (2014). Questioning the theory fetish in information systems. *Journal of Information Technology*, 29(4), 327-336.
- Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483, 531-533.
- Berthon, P., Pitt, L., Ewing, M., & Carr, C. L. (2002). Potential research space in MIS: A framework for envisioning and evaluating research replication, extension, and generation. *Information Systems Research*, 13(4), 416-427.
- Campbell, D. T. (1986). Science's social system of validity-enhancing collective belief change and the problems of the social sciences. In Fiske, D. W. & Schweder, R. A. (Eds.), *Metatheory in social science: Pluralisms and subjectivities* (pp. 108-135). Chicago: University of Chicago Press.
- Carey, B. (2015). Many psychology findings not as strong as claimed, study says, *New York Times*. Retrieved from <http://nyti.ms/1JoE6N6>.
- Chen, L., Po-An Hsieh, J. J., Van de Vliert, E., & Huang, X. (2015). Cross-national differences in individual knowledge-seeking patterns: A climato-economic contextualization. *European Journal of Information Systems*, 24(3), 314-336.
- Collberg, C., & Proebsting, T. (2016). Repeatability in computer systems research. *Communication of the ACM*, 9(3), 62-69.
- Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice*. Beverly Hills, CA: Sage.
- Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. *Academy of Management Journal*, 50(6), 1281-1303.
- Dearden, J. (1972). MIS is a mirage. *Harvard Business Review*, 50(1), 90-99.
- Dennis, A. R., & Valacich, J. S. (2014). A replication manifesto. *AIS Transactions on Replication Research*, 1, 1-4.
- Edmondson, A. C. (2011). Strategies for learning from failure. *Harvard Business Review*, 89(4).
- Eisend, M., & Tarrahi, F. (2014). Meta-analysis selection bias in marketing research. *International Journal of Research in Marketing*, 31(3), 317-326.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4(5), e5738.
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904.

- Frank, U. (2006). Towards a pluralistic conception of research methods in information systems research /*ICB-Research Reports* (Vol. No. 7, pp. 84). Essen, Germany: Institut für Informatik und Wirtschaftsinformatik (ICB).
- Gad-El-Hak, M. (2004). Publish or perish—An ailing enterprise? *Physics Today*, 57(3), 61-62.
- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science*, 351(6277).
- Gregor, S. (2014). Theory – Still king but needing a revolution! *Journal of Information Technology*, 29(4), 337-340.
- Gregor, S., & Klein, G. (2014). Eight obstacles to overcome in the theory testing centre. *Journal of the Association of Information Systems*, 15(11), i-xix.
- Grover, V., & Lyytinen, K. (2015). New state of play in information systems research: The push to the edges. *MIS Quarterly*, 39(2), 271-296.
- Hempel, C. (1966). *Philosophy of natural science* NJ: Englewood Cliffs, Prentice-Hall.
- Hubbard, R., & Armstrong, J. S. (1994). Replication and extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing*, 11, 233-248.
- Hunter, S. D. (2010). Same technology, different outcome? Reinterpreting Barley's. Technology as an Occasion for Structuring. *European Journal of Information Systems*, 19(6), 689-703.
- Kahneman, D. (2012). Open Letter Retrieved 2014-04-28, from  
<http://www.decisionsciencenews.com/2012/10/05/kahneman-on-the-storm-of-doubts-surrounding-social-priming-research/>.
- Kane, E. J. (1984). Why journal editors should encourage the replication of applied econometric research. *Quarterly Journal of Business Economics*, 23(Winter), 3-8.
- KC, D., Staats, B. R., & Gino, F. (2013). Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery *Management Science*, 59(11), 2435-2449.
- Koufaris, M. (2010). Commentary on Hunter's 'Same technology, different outcome? *European Journal of Information Systems*, 19(5), 704-706.
- Keen, P. G. W. (1980). *MIS research: Reference disciplines and a cumulative tradition*. Paper presented at the International Conference on Information Systems, Philadelphia, PA.
- Lee, A., & Baskerville, R. (2003). Generalizing generalizability in information systems research. *Information Systems Research*, 14(3), 221-243.
- Lee, A., & Baskerville, R. (2012). Conceptualizing generalizability: New contributions and a reply. *MIS Quarterly*, 36(3), 749-761.
- Leibniz, G. (1765). *Nouveaux Essais sur l'entendement humain*. Amsterdam, Leipzig: Raspe RE.
- Lyytinen, K., & King, J. L. (2004). Nothing at the center? – Academic legitimacy in the information systems field. *Journal of the Association for Information Systems*, 5(6), 220-246.
- Markus, L. M. (2014). Maybe not the king, but an invaluable subordinate: A commentary on Avison and Malaurant's advocacy of 'theory light' IS research. *Journal of Information Technology*, 29(4), 341-345.
- Maturana, H. R. (1987). Kognition. In S. J. Schmidt (Ed.), *Der Diskurs des radikalen Konstruktivismus* (1st ed., pp. 89-118). Frankfurt am Main: Suhrkamp.
- McCloskey, D. N. (1985). *The rhetoric of economics*: Madison.
- Nansen, B., Arnold, M., Wilken, R., & Gibbs, M. (2012). Broadbanding Brunswick – High-speed broadband and household media ecologies: A report on household take-up and adoption of the national broadband network in a first release site. In A. C. C. A. Network (Ed.). Sydney
- OECD. (2013). OECD factbook 2013: Economic, environmental and social statistics. Retrieved 2014-05-01, from <http://www.oecd.org/sti/msti.htm>.

- Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York: Routledge.
- Popper, K. (1980). *The logic of scientific discovery*. London: Unwin Hyman.
- Preston, D., Karahanna, E., & Rowe, F. (2006). Development of shared understanding between the chief information officer and top management team in U.S. and French organizations: A cross-cultural comparison. *IEEE Transactions on Engineering Management*, 53(2), 191-206.
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(712).
- Rorty, R. (1999). *Philosophy and social hope*. New York: Penguin Books.
- Rowe, F. (2011). Towards a greater diversity in writing styles, argumentative strategies and genre of manuscripts. *European Journal of Information Systems*, 20(5), 491-495.
- Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations. *European Journal of Information Systems*, 23(3), 241-255.
- Salovaara, A., & Merikivi, J. (2015). *IS Research progress would benefit from increased falsification of existing theories*. Paper presented at the European Conference of Information Systems, Münster.
- Sayer, A. R. (1992). *Method in social science* (2nd ed.). London, UK: Routledge.
- Seddon, P., & Scheepers, R. (2012). Towards the improved treatment of generalization of knowledge claims in IS research: Drawing general conclusions from samples. *European Journal of Information Systems*, 21(1), 6-21.
- Seddon, P., & Scheepers, R. (2015). Generalization in IS research: A critique of the conflicting positions of Lee & Baskerville and Tsang & Williams. *Journal of Information Technology*, 30(1), 30-44.
- Shadish, W., Cook, T., & Campbell. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Spagnoletti, P., Za, S., & Winter, R. (2013). *Exploring foundations for using simulations in IS research*. Paper presented at the International Conference on Information Systems, Milano.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: University Press.
- Tsang, E., & Williams, J. (2012). Generalization and induction: Misconceptions, clarifications, and a classification of induction. *MIS Quarterly*, 36(3), 729-748.
- Tsang, E. W. K., & Kwan, K.-M. (1999). Replication and theory development in organizational science: A critical realist perspective. *The Academy of Management Review*, 24(4), 759-780.
- Unreliable research – Trouble at the lab. (2013, October 19). *The Economist*, 26-30.
- Warkentin, M., Charles-Pauvers, B., & Chau, P. (2015). Cross-cultural IS research: Perspectives from Eastern and Western traditions. *European Journal of Information Systems*, 24(3), 229-233.
- Wittgenstein, L. (1961). *Tractatus logico-philosophicus*. London, New York: Routledge & Pau.
- Weber, R. (1997). *Ontological foundations of information systems* (1. ed.). Melbourne, Australia: Coopers & Lybrand.
- Yong, E. (2012). Replication studies: Bad copy. *Nature*, 485 (17 May 2012), 298-300.

## About the Authors

**Sebastian Olbrich** is Professor of Information Systems and Digital Business at the European Business School (EBS) in Wiesbaden, Germany. He also has a long experience as a Strategy Consultant and an Interim-Manager in the field of Information Systems. Next to his work as a practitioner, he published more than 50 peer-reviewed articles, serves for IS conference and Journals in various roles and lectures in multiple institutions such as the University of Mannheim, the European Business School or University of Mainz. His research interest is the knowledge transfer between theory and practice. Hence, he works on making the practical impact of research explicit, works in executive education and is speaker at various IS and CPM related conferences such as TDWI, Management Circle, European Finance Week, IBM Business Connect or CeBIT.

**Ulrich Frank** Ulrich Frank holds the chair of Information Systems and Enterprise Modelling at the Institute of Computer Science and Business Information Systems at the University of Duisburg-Essen. His main research topic is enterprise modelling, i.e. the development and evaluation of modelling languages, methods and corresponding tools. Further areas of research include method construction, models at run time, methods for IT management and research methods. Ulrich Frank is on the editorial board of the journals "Enterprise Modelling and Information Systems Architectures", "Business & Information Systems Engineering", "Software and Systems Modeling", "Information Systems and E-Business Management", and the "Journal of Information System Modeling and Design". He has been actively involved in numerous conferences and various major research projects. He had assignments as visiting researcher/professor in various countries. Ulrich Frank is the founding director of the international student exchange network IS:link.

**Shirley Gregor** is Professor of Information Systems at the Australian National University, Canberra, and Associate Dean Research in the College of Business and Economics. Her research interests include the innovative use of information technology, knowledge systems and the philosophy of technology. Dr Gregor has led several large applied research projects funded by government agencies, the Australian Research Council and AusAID. Her publications appear in journals including MIS Quarterly, Journal of Management Information Systems, Journal of the Association of Information Systems, European Journal of Information Systems and Information Technology & People. She was Editor-in-Chief of the Journal of the Association for Information Systems 2010-2013. She has an abiding interest in epistemological issues, particularly the development of sound bodies of knowledge and theory in a practical discipline such as information systems.

**Fred Niederman** serves as Shaughnessy Professor of MIS at Saint Louis University. He teaches project management, systems analysis, database, and research methods. He has published on the nature of theory and is working on additional studies in this area. He has published in MIS Quarterly, Journal of Strategic Information Systems, JAIS, Journal of Information Technology, Information Systems Journal, CACM, Journal of Management Information Systems, and the Journal of Global Information Management. He is currently servicing as coordinator of the senior scholars which includes presenting an annual forum at ICIS. He serves as department co-editor for IS for the Project Management Journal, senior editor for JAIS and CAIS, and editorial roles for a variety of other IS and management journals.

**Frantz Rowe** serves as a Professor of Information Systems at the University of Nantes, France. He was trained as an Engineer (Master in Civil Engineering) and as an Economist (Master of Arts). His PhD from the University of Paris was awarded in 1991. He has been an invited Professor at Harvard University, Bentley University and Cape Town University. His current research areas pertain to organizational transformations related to information systems projects and to the effects of PLM use. He has recently directed several research projects on ERP dynamics, on the design and performance of call centers and on electronic marketplaces and supply chains. He has published 61 articles in scientific journals and several books, including Innovation and IT in an international context (2014, with Dov Te'eni). He has been the Editor in Chief of Systèmes d'Information et Management and of the European Journal of Information Systems. He has directed 25 PhD students and is currently directing the PhD program in Management.

Copyright © 2016 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from [ais@aisnet.org](mailto:ais@aisnet.org).