

Simulation-Based Research in Information Systems: Epistemic Implications and a Review of the Status Quo

Abstract: *Simulations provide a useful methodological approach for studying the behavior of complex socio-technical information systems (IS), in which humans and IT artefacts interact to process information. However, the use of simulations is relatively new in IS research and the current presence and impact of simulation-based studies is still limited. Furthermore, simulation-based research is quite different from other approaches, making it difficult to position and evaluate adequately. Therefore, this paper first analyses the particularities of simulation-based IS research that distinguish this approach from other, more conventional approaches. Based on this analysis, we then conduct a structured literature review of the status quo of simulation-based IS research, to understand how IS researchers currently employ simulation. In combination, a comparison of the theoretical potential of simulation-based research in IS with its current status quo enables us to derive a set of propositions, which provide guidance for prospective simulation-based research.*

Keywords: Simulation, Epistemology, Socio-technical system, Literature review

Jannis Beese

Institute of Information Management
University of St. Gallen
Unterer Graben 21, 9000 St. Gallen, Switzerland
jannis.beese@unisg.ch

Dr. Mohammad Kazem Haki

Institute of Information Management
University of St. Gallen
Unterer Graben 21, 9000 St. Gallen, Switzerland
kazem.haki@unisg.ch

Prof. Dr. Stephan Aier

Institute of Information Management
University of St. Gallen
Unterer Graben 21, 9000 St. Gallen, Switzerland
stephan.aier@unisg.ch

Prof. Dr. Robert Winter

Institute of Information Management
University of St. Gallen
Unterer Graben 21, 9000 St. Gallen, Switzerland
robert.winter@unisg.ch

1 INTRODUCTION

Simulation, which is broadly defined as “a method for using computer software to model the operation of real-world processes, systems, or events” (Davis et al. 2007, p. 481), provides a distinctive methodological approach for studying various phenomena in different disciplines. The use of simulation allows researchers to isolate and vary the potentially large number of parameters of the respective system in a controlled environment, while producing massive amounts of data that, for instance, enable researchers to capture non-linear relations with statistical techniques. Thus, scholars have recently reiterated the potential of computer simulations to advance the information systems (IS) discipline by providing a novel way to investigate IS phenomena (Burton and Obel 2011; Loos et al. 2013; Spagnoletti et al. 2013; Zhang and Gable 2014).

While simulation-based research has substantially contributed to other disciplines, for instance to the natural sciences or computer science, its current presence and impact in IS research is comparatively low (Zhang and Gable 2014). This can be due to (i) the nature and peculiar character of IS phenomena as well as to (ii) particularities of simulation-based research approaches, both subject to closer investigation. *First*, the social and the technical aspects of IS phenomena are inextricably intertwined so that boundaries between these two aspects are not clear-cut (Bostrom and Heinen 1977; Lyytinen and Newman 2008). This entails a need to conceptualize IS phenomena as *complex socio-technical* ensembles (Luna-Reyes et al. 2005; McLeod and Doolin 2012), for which both the social and the technical aspects and their relationships should be taken into account. *Second*, compared to theoretical analysis or deduction as well as empirical analysis or induction, simulation is recognized as a distinct third way of doing science (Harrison et al. 2007). Using simulation techniques, analytical reflections can be captured through mathematical models, which provide their own virtual data to overcome the problem of data availability in empirical investigations (Harrison et al. 2007). As such, scholars try to model and translate real-world problems into a virtual (i.e., simulation) world with the aim to derive meaningful insights about the real-world problems. Therefore, the creation of scientific knowledge through simulation-based research requires several *epistemic* (how knowledge can be justified through simulation results) and *methodological* (regarding the activity of simulating and its relation to theorizing and experimenting) considerations (Rohrlich 1990; Dowling 1999; Grüne-Yanoff and Weirich 2010). These considerations are decisive in IS research due to complexity of the underlying social systems as well as uncertainties inherent in incompletely described technical artefacts (Curşeu 2006).

This study first aims to understand how simulation-based research approaches can be used to study complex socio-technical IS phenomena. We adopt a socio-technical perspective to explain IS phenomena and discuss the epistemic inferences that are made in the simulation-based research process (Sargent 2005) of examining IS phenomena. Subsequently, we use the discussed epistemic inferences as a foundation to conduct a systematic literature review on simulation-based research in IS¹ in order to understand how simulation-based approaches are currently employed in IS research. Finally, the results of both the epistemological implications of simulation-based research and the literature review are discussed and contrasted, allowing us to derive a set of propositions that provide guidance for prospective simulation-based research in IS.

2 STUDYING IS PHENOMENA THROUGH SIMULATION

To understand the challenges of studying IS phenomena through simulation-based approaches it is necessary to discuss the particularities of IS phenomena. In their investigation of the “intellectual core” of the IS discipline, Sidorova et al. (2008) find that contemporary IS research generally includes the social context in

¹ A part of this literature review was previously presented at (blinded for peer review). In the paper at hand we significantly extend the literature review and focus on the additional insights that follow from the discussion of the epistemic particularities of simulation-based IS research.

which technical artefacts are designed and used. That is, while IS research essentially studies human-made technology-based systems, this analysis requires an understanding of corresponding social systems as well as the interaction between the social and technical systems (Lee 2010; Becker et al. 2015). This is what Orlikowski (1992) calls “duality of technology”—the dialectical interaction between technical artifacts and their social context—and what Orlikowski and Iacono (2001) call “ensemble view of technology”—dynamic interactions between technical artifacts and their social context.

Consequently, our conceptualization of IS phenomena adopts a socio-technical systems perspective to identify and organize the constituent parts of IS (Bostrom and Heinen 1977; Lyytinen and Newman 2008; Wu et al. 2015). This perspective conceptualizes IS as two interrelated subsystems, the *technical system* and the *social system*. The technical system is concerned with the processes, tasks, and technologies that are needed to acquire, store, and transform information to outputs, such as products or services. The social system is concerned with the relationships among people and the attributes of these people, such as attitudes, skills, and values. Recent contributions to the socio-technical perspective conceptualize IS as a mutually interactive socio-technical ensemble comprising *actors*, *tasks*, *technology*, and *structure*, reflecting both technical and social aspects and the relations among them, which are embedded in and influenced by an external *environment* (Lyytinen and Newman 2008; Gregoriades and Sutcliffe 2008; Wu et al. 2015). Table 1 gives an overview of the constituent components of IS as socio-technical systems.

Component	Exemplary elements and properties
Tasks	Goals; project deliverables and time requirements; business requirements and supported processes; stakeholder expectations
Actors	All relevant stakeholders, e.g., developers, workers, administrators, and their personal properties; values and beliefs
Structure	Communication systems and structures both formal, e.g., project meetings, and informal, e.g., break discussions; organizational hierarchies; workflows
Technology	Tools; work systems; technical infrastructure; software and hardware technology
Environment	External influences on the system, such as regulatory requirements and market-driven changes

Table 1: Constituent components of IS as socio-technical systems (Lyytinen and Newman 2008)

What makes socio-technical systems distinct and subject to closer investigation is the overall behavior of such systems, which is dependent on a diverse set of often non-linear and dynamic mechanisms that relate to both the social and technical subsystems (Luna-Reyes et al. 2005; McLeod and Doolin 2012). Lyytinen and Newman (2008) argue that changes in IS, as socio-technical systems, occur due to misalignments among the systems’ constituent socio-technical components (as shown in Table 1). They posit that change is not solely or even mainly incremental and cumulative, but rather is episodic and punctuated. As such, along with evolutionary (first-order) changes to fix the misalignment among components, a socio-technical system experiences revolutionary (second-order) changes (i.e., punctuations) over time by which components of the system become re-configured, and the given system eventually exhibits new, emergent properties. Due to the complexity of socio-technical systems and their non-deterministic behavior, studying such systems through conventional research methods (e.g., case study, survey) is challenging and highly effort-consuming. Simulation-based research has been proposed as a suitable alternative for studying complex socio-technical systems, since simulations can be performed with less resources and in a controllable fashion (Curşeu 2006).

2.1 Epistemic Inferences of Simulation

While simulation is useful for studying socio-technical systems, the use of simulation entails a careful reflection of real-world problems as well as of existing understandings of the real-world problems in

simulation models. It also requires an accurate evaluation and validation of simulation experiments and a thorough interpretation of the resulted insights' implications to the given real-world problems. In simulation-based studies, scholars translate real-world problems into a virtual (i.e., simulation) world with the aim of deriving meaningful insights for the real-world problems, which demands careful reflection and interpretation. Each step of a simulation process can be questioned in terms of how appropriate the reflection of a real-world problem in the simulation model is (transferring the existing real-world knowledge to the simulation world), how reliable the simulation experiments are (simulation), and how scientific and how meaningful the resulted insights of the simulation to the given real-world problems are (transferring resultant knowledge from the simulation and making sense of this knowledge in the real-world context). To elaborate on these inherent challenges of simulation-based research, we first discuss different simulation techniques, then outline the constituent steps of a simulation-based research process and eventually discuss epistemic inferences in different steps of such a process.

The term simulation in our adopted definition by Davis et al. (2007) refers to a very diverse bunch of methods, each with its own epistemic capabilities and restrictions. Hence, we distinguish different simulation techniques in our analysis, namely, *analytical simulations*, *stochastic processes*, *system dynamics*, *genetic algorithms*, *artificial neural networks*, and general *agent-based* simulations, in line with previous literature on simulation based research in IS (e.g., Davis et al. 2007; Spagnoletti et al. 2013; Zhang and Gable 2014). *Analytical simulation* refers to simulation models that are directly based on a mathematical description of a system and which focus the use of formal models (e.g., game theory, and auctioning theory). *Stochastic processes* are similar regarding the formal nature of the underlying mathematical models, but differ in that at the center of such simulations is a model of a stochastic process, for example characterized by a random walk (Ripley 2009). Such simulations are frequently employed to study the consequences of changes to highly stylized socio-economic market models (e.g., Xiao and Dong 2015). *System dynamics* refers to a specific modelling approach that employs the concept of stocks and flows to simulate the behavior of complex systems over time (Roberts et al. 1994; Sterman 2002). *Genetic algorithms* refers to an optimization approach with roots in biology that models a system as a set of heterogeneous entities—called candidate solutions—which then iteratively evolve and adapt over time to better match a given fitness function (Whitley 1994; Davis et al. 2007). Similarly, *artificial neural network* refers to a specific machine learning technique that is commonly employed in simulations as a component that adapts to and learns from input data (Graupe 2013). Finally, we distinguish other, general *agent-based* models, referring to any computational model that represents the actions and interactions of autonomous agents (Macal and North 2009).

Following Sargent (2005)(2005)(2005), the simulation model development process comprises distinct steps, starting with the selection of the real-world system to be simulated (*problem entity*) and ending with the actual implementation of the *simulation model* on a computer (see the upper part of Figure 1). Researchers can then employ the implemented simulation model to conduct simulation experiments and use the obtained results in combination with empirical data or extant system theories to create new knowledge (lower part of Figure 1).

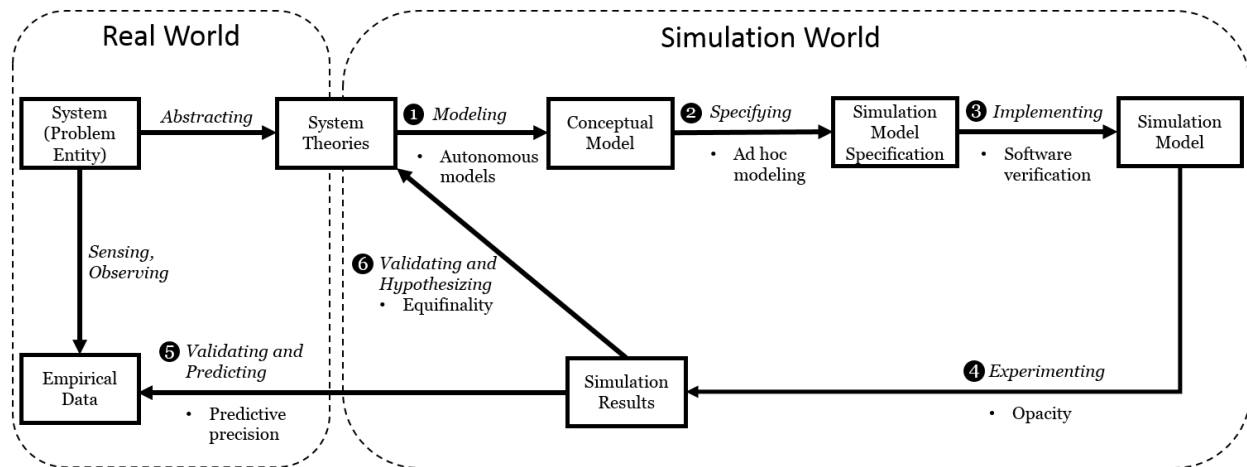


Figure 1: Constituent steps of a simulation-based research process

The term *problem entity* is used to refer to the real-world *system* in which the investigated phenomenon of interest is situated. Researchers generally rely on scientific theory (*system theories*), which describes extant knowledge about the problem entity, to derive a *conceptual model* as a foundation for the design of the simulation. While this model of Sargent (2005)(2005)(2005) is quite general, this ensures that it is applicable to a wide range of simulation-based research, making it a useful starting point for our subsequent analysis. Furthermore, it has been widely used (Law 2008; Bratley et al. 2011) and other simulation process models may use slightly different terminology, but are very similar in their structure (e.g., Davis et al. 2007).

The relation between models and theory is the subject of an ongoing debate in the philosophy of science (Frigg and Hartmann 2012; Bichler et al. 2016). A common position is that models are independent from theory in their construction and their functioning (Winsberg 1999; Morgan and Morrison 1999; Grüne-Yanoff and Weirich 2010). In the context of simulation-based research this means that scholars are required to interpret and refine scientific theory, which is rarely precise enough to directly translate into computerized simulation models. Along this process, one can distinguish different types of models that are used in simulation-based research (Winsberg 2003; Sargent 2005; Küppers and Lenhard 2005). We use the term *conceptual model* (Sargent 2005)(Sargent 2005)(Sargent 2005) to refer to the basic conceptual understanding of the problem entity that scholars gain through interpreting respective scientific theory. Winsberg (2003) refers to this as a principle model: “The simulationist begins by choosing a principle model—a model that characterizes the system in terms of both the arrangement of its constituent parts, and the rules of evolution that govern the changes of state that the system undergoes over time” (Winsberg 2003, p. 108).

The term *simulation model specification* refers to the mathematical model that represents and reflects the conceptual model in the simulation. Since any computer implementation of a simulation necessitates a precise mathematical definition of all employed constructs and relations, researchers have to make simplifications and assumptions to fill the parts of the conceptual model that lack the required mathematical precision. This simplification, which brings about specific assumptions about the simulation model, should account for stylized facts that only include abstract and relevant aspects of the real world in the simulation model (Bichler et al. 2016). Finally, we use the term *simulation model* to refer to the executable simulation, i.e., the actual implementation on a computer.

Scientific simulations that represent socio-technical systems are not simply calculations, but instead “involve a complex chain of inferences that serve to transform theoretical structures into specific concrete knowledge” (Winsberg 1999, p. 275). This process of knowledge-generation through simulation is described in Figure 1, following Sargent (2005). We focus our analysis on the links between knowledge of the real-world system (in the form of scientific system theories and empirical data) and the simulation world

(steps 1, 5, and 6 in Figure 1), as well as the links between two intermediate constructs within the simulation world (steps 2, 3, and 4 in Figure 1). Each of these creates another layer of distance between reality and the obtained simulation results, thus requiring epistemic justification.

Conceptual Modeling (1): Since “any reasonably comprehensive simulation of organizations must be constructed from insights made with regard to how organizations have been observed to operate,” (Kulik and Baker 2008, p. 88) we analyze the construction of conceptual models based on scientific theory. The importance of a theory-informed conceptualization is underlined by Sargent’s choice of including the abstraction of a system theory as the basis of the conceptual model as a fundamental step in the simulation model development process (Sargent 2005). Here, system theory refers to any kind of theory from which the conceptual model can be derived such as IS modeling theories, which are already discussed in IS modeling literature (Bichler et al. 2016). While a unified modeling theory is still missing, depending on the targeted conceptual model, researchers can employ a wide range of theories such as fuzzy set theory or alternative uncertainty theories, including stochastics (Bichler et al. 2016).

The construction of conceptual models in simulation-based research is “an activity that often brings us beyond the original theoretical principles themselves” (Winsberg 2003, p. 118), since conceptual modeling usually involves creativity and intuition (Winsberg 2003; Frank and Troitzsch 2005). Philosophers of science often use the term *autonomous models* (Morgan and Morrison 1999), in the sense that models “function as *instruments* of investigation, [as they] are partially independent of both theories and the world” (Morgan and Morrison 1999, p. 10). Following Winsberg (2003), the view “that models are autonomous or independent of theory is meant to emphasize the fact that there is no algorithm for reading models of theory” (Winsberg 2003, p. 106). On one hand, the autonomy of models therefore requires researchers to argue that the creative, intuitive modeling choices lead to an accurate description of the investigated phenomenon in the results obtained from the final simulation model (in step 4). On the other hand, it is often this very process that brings about new scientific knowledge, for example through providing additional evidence for or against intuitive refinements of extant theory. Researchers are thus often required to revisit this step after comparing the resulted insights of the simulation model with extant theories or empirical studies, which makes simulation-based research an iterative process (Sargent 2005).

Specifying (2): The act of *specifying* concerns the development of a mathematical representation of the conceptual model. Similarly to the previous step, researchers are required to make assumptions in translating a conceptual model to a computational model (Poile and Safayeni 2016), which is a transformations process that generally involves “idealizations, approximations, and even self-conscious falsifications” (Winsberg 2003, p. 108). As implemented simulation models correspond to precise mathematical constructs, a preliminary step in their creation requires that “specific rules replace general laws” (Pias 2011, p. 35). Winsberg (2003) refers to this creative process as *ad hoc modeling*: “Ad hoc modeling includes such techniques as simplifying assumptions, removal of degrees of freedom, and even substitution of simpler empirical relationships for more complex, but also more theoretically-founded laws” (Winsberg 2003, p. 109). As such, the resultant simulation model often only reflects stylized facts, i.e., interesting patterns in empirical data that focus statistical relations between observable phenomena while abstracting from details (Houy et al. 2015). This endeavor simplifies empirical data and makes specific assumptions about the simulation model explicit. However, due to the imprecisions introduced by ad hoc modeling, the simulation model does not have the full faith and credit of the governing theory’s epistemic credentials and thus requires additional justification and validation. One of the promising ways to ensure the validity of the simulation models is to employ stylized facts based on existing system theories (Houy et al. 2015; Bichler et al. 2016). By doing so, the resultant simulation model comprises both scientific theory and real-world empirical data, essentially mediating between theory and real-world systems (Bichler et al. 2016).

Implementing (3): From an epistemic perspective, the implementation of a specified model on a computer requires a researcher to argue that the simulation model is an error free and correct software implementation of the simulation model specification on the given physical hardware. A formal, algorithmic code verification is often not feasible for complex software (D’Silva et al. 2008; Ashish and Aghav 2013).

Instead, researchers usually argue that the process of implementing the simulation model was rigorous, for instance by adhering to commonly recommended processes and software engineering principles.

Experimenting (4): The central epistemic issue when conducting simulation experiments is related to the inability of researchers to analytically trace how the simulation results were obtained. This is often referred to as epistemic *opacity*: it is simply impossible for humans to follow and understand the millions of calculations that are performed by the computer to obtain the results of simulation experiments (Humphreys 1990; Grüne-Yanoff and Weirich 2010). This is particularly true and even more decisive for simulations that aim to understand emergent phenomena, a common goal in the exploration of socio-technical systems (Kochanowicz et al. 2013). The idea of *emergence* in socio-technical simulation experiments has been nicely described by Coleman (1994): to understand macro-level associations between observed macro-level socio-technical phenomena, one investigates how the occurrence of a given macro-level phenomenon affects individual elements of the analyzed socio-technical system and how these individual elements interact and influence each other. In consequence, these interactions again aggregate to the observed macro-level phenomena (Coleman 1994; Boero and Squazzoni 2005; Manzo 2007; Hedström and Ylikoski 2010).

As it is not possible for a human to follow all the calculations that model these situational, action-formation, and transformational mechanisms, one has to instead rely on additional techniques, such as graphical visualization, to argue for the relation between simulation design and observed output data from simulation experiments (Trier 2008; Lee et al. 2015). IS researchers may rely on the extant knowledge from related disciplines, such as research on the design of autonomous software agents and multi-agent simulations (Birdsey and Szabo 2014; Doan et al. 2014), to structure their simulation experiments.

Validating and Predicting (5): Validation of simulation results can either be done by comparing them with data obtained from empirical studies (this step, *validating and predicting*, in Figure 1) or by evaluating them in the context of already established scientific theory (step 6, *validating and hypothesizing*, in Figure 1). Using the first approach, the epistemic credit of simulation results is largely related to the ability of the simulation to reproduce or predict characteristics of the socio-technical phenomenon under investigation (Boero and Squazzoni 2005). Consequently, a large number of validation techniques relies on a direct comparison with real-world observations to argue that simulation results constitute scientific knowledge.

In addition to using empirical data for the validation of simulation results, researchers also develop simulations with the goal to predict future empirical data. Both, testing a simulation's *predictive validity* and using a simulation to predict, is complicated through epistemic issues related to the *predictive precision* of a simulation model. In general, due to abstractions and simplifications in the simulation model development process, the results of simulation experiments do not show a precise one-to-one correspondence with empirical data. Instead, simulation experiments rather aim to create statistical estimates of phenomena or observations that suggest the presence of certain conceptual relations (Küppers and Lenhard 2005). Thus, scholars have argued that in the empirical validation of simulation results “adequacy replaces proof” (Pias 2011, p. 35), meaning that no direct relation to reality is established, but simulation results are instead judged against experience and high-level observations. Similarly, “performance beats theoretical accuracy” (Küppers and Lenhard 2005, p. 6) when using simulations to predict, meaning that simulation models are rather evaluated against their utility than against a perfect correspondence with reality. In consequence, however, this lack of *predictive precision* requires researchers to be very careful in framing and positioning their results from an epistemic perspective.

Validating and Hypothesizing (6): In addition to relying on empirical observations, validation of simulation results can also be done by comparing them with established scientific theory. To that end, simulation results are contrasted with a theory-informed understanding of the phenomenon of interest, in order to test if the simulation is able to adequately recreate extant scientific knowledge. Respective validation techniques aim to facilitate this process of data generation and interpretation, for example by describing common types of visualizations or specific parameterization strategies.

Furthermore, simulation results can also be used to advance and refine extant knowledge, for example through evaluating the influence of certain parameterization choices and modeling decisions that were made in simulation development to derive novel hypotheses and propositions in a given theoretical framework (Epstein 2008). Contrary to the traditional epistemic process commonly employed in natural sciences, epistemological inference is “downwards” in simulation-based research, starting with abstract theory and then going to empirical observations (Winsberg 2001). This leads to the issue of *equifinality*: different initial conditions and different processes can lead to the same final result, leading to a potential problem in deducing novel theory from complex simulations (Epstein 1999; Harrison et al. 2007; Davis et al. 2007; Weinhardt and Vancouver 2012; Poile and Safayeni 2016). How can one argue about theoretical constructs and causal processes based on the observation that “two black boxes are able to reach the same outcome” (Poile and Safayeni 2016, p. 4)? Researchers are therefore usually required to employ multiple validation techniques to not only understand structure in the simulation results, but also how these results were generated (Sargent 2005). In addition to discussions within the IS discipline on emergence in complex socio-technical systems (Lyytinen and Newman 2008; Lee et al. 2015; Prat et al. 2015), related research on the verification of autonomous software agents and multi-agent simulations proposes several techniques and frameworks (e.g., Doan et al. 2014; Montali et al. 2014; Aminof et al. 2016; Jamroga et al. 2016) deal with issues related to this black box nature of simulation experiments.

3 LITERATURE SELECTION AND CODING PROCEDURE

Relying on the preceding discussion of epistemological particularities of simulation-based IS research, we now investigate how IS scholars conduct such research. To assess the status quo of extant simulation-based research in IS, we opt for a systematic literature review, following the suggestions of Webster and Watson (2002) and vom Brocke et al. (2015).

3.1 Literature Selection

To make our review of the pertinent literature as comprehensive as possible, we opt for a literature selection procedure that determines a representative set of papers from the large body of related publications (vom Brocke et al. 2015). We therefore adopt the list of 21 top IS journals analyzed by Lowry et al. (2013) as well as conference papers presented at both of the most influential international and European conferences in IS (i.e., ICIS and ECIS). To identify relevant publications, we conducted a search via the ISI Web of Science using the different simulation techniques introduced in section 2.1, as well as the general term *Simulation*. In addition, we searched in the AIS electronic Library (AISEL) for papers presented at ECIS or ICIS. We used the following search string for the fields abstract, title, and keywords, across all selected journals for publications published up to the year 2016:

“neural network” OR “system dynamic” OR “NK fitness landscape” OR “genetic algorithm*” OR “cellular automat*” OR “stochastic process*” OR “simulation*”*

In total, we retrieved 697 publications in this first step. To select the most relevant and influential papers from this database, we first only included papers that have on average at least one citation per year since their publication. However, as papers published in 2015/2016 are too recent to get a large number of citations, and as papers published in AIS’s basket of top journals are crucial for our review, we included all of them in our review. This results in 255 papers, for which we then read the abstract and introduction sections, to exclude papers that do not develop an own computer simulation model. In this step, we excluded, for example, papers that only discuss the use of simulation in general, only reference other simulation studies, or use simulation to refer to human experiments that do not involve computation. The final result is a database of 175 relevant papers (see Appendix A1).

3.2 Analysis Framework and Coding Procedure

For coding the selected papers, we developed a comprehensive analysis framework, based on the preceding discussion of epistemic inferences. We thereby focus on the links between knowledge of the real-world

system, in the form of scientific system theories or empirical data, and the simulation world. We distinguish three parts to form the analysis framework and to structure the subsequent discussion of results:

Real world to simulation world: First, we investigated how researchers went about constructing conceptual models (*step 1, modelling* in Figure 1). Since scientific knowledge about the real-world systems is essentially based on scientific theories (Bichler et al. 2016), we in particular analyzed how researchers employed extant system theories in both the design of conceptual simulation models and the interpretation of simulation results (Sargent 2005; Davis et al. 2007; Poile and Safayeni 2016). In coding our results, we relied on the list of 174 theories used in IS research, which are retrieved by Lim et al. (2013). Due to the large number of theories and the large number of papers in our database, we derived regular expressions for each theory (e.g. “*Resource dependence theory*” was replaced by “*resource.?dependenc*” to capture instances where “*resource-dependency*” is written with a hyphen and a “y”), which were then used in combination with text-mining software to color-code relevant parts of the analyzed papers. Then we read the respective papers, to see how theories are used, and only included a theory in our coding if it is explicitly referenced and employed in constructing the simulation model or in evaluating simulation results.

Simulation world: Second, we analyzed how researchers translated their conceptual model to a simulation model, relating to the steps *specifying (step 2)* and *implementing (step 3)* in Figure 1. Based on the preceding discussion of IS as socio-technical systems (see Section 2), we classify simulation-based research according to the socio-technical system components (i.e., *actors, tasks, structure, technology, and environment*) that were included in the simulation model, to understand which aspects of the investigated IS were focused in the model and where potential simplifications may take place. We additionally coded papers according to the employed simulation technique (*stochastic processes, analytical, system dynamics, genetic algorithms, artificial neural networks, and agent-based*), based on the discussion in section 2.1. Coding mainly relied on reading the content of the methodology sections of the selected papers, reaching into other sections whenever the model was discussed in a different part.

Simulation world to real world: Third, we studied how the simulation was used to investigate and validate against real-world phenomena, combining the steps *experimenting (step 4)*, *validating and hypothesizing (step 5)* and *validating and predicting (step 6)*. Since simulations may fulfill a very different purpose in scientific research (Boero and Squazzoni 2005), we coded simulations by their intended use, as this significantly affects the development process of the simulation model (Davis et al. 2007). Following Harrison et al. (2007) and Axelrod (1997), we distinguish the following uses for simulation: *prediction* (how are model variables related), *proof* (show that certain system behavior exists), *discovery* (discover unexpected consequences of interactions), *explanation* (explain why the system behaves in a certain way), *critique* (test existing theories), *prescription* (suggest how to best interact with or within the system), and *empirical guidance* (derive hypotheses for empirical testing). During coding, we relied on the descriptions given by Harrison et al. (2007, pp. 1238-39) to investigate the introduction, results, and discussion sections of the selected papers. Note that it is possible for a simulation to serve multiple purposes. For example, Wöhner et al. (2015) use a simulation model to first *predict* the effects of different parametrizations on the behavior of managed wikis, and subsequently employs the obtained data to *prescriptively* argue how this concept can be used to overcome related issues, such as online harassment and cyberbullying (Wöhner et al. 2015).

Finally, we rely on Sargent (2005) to distinguish different validation techniques that researchers may employ in the process of simulation-based research (see Table 2). For coding, we read the methodology, results, and discussion sections of the selected papers, as well as related appendices.

Validation technique	Description
Predictive validity	Use the model to make predictions and compare them with real-world behavior
Historical data validation	Use historical data not only for calibration but also for testing, e.g., by splitting the dataset in half
Event validity	Compare simulated events with those occurring in real-world systems

Turing tests	Check if experts can discriminate between model and real-world system output
Animation	Display operational behavior graphically, e.g., show the movements of actors
Comparison with other models	Compare the model with other extant validated models
Extreme condition tests	Check if the model behaves reasonably when extreme values are selected for model parameters
Degenerate tests	Test degeneracy of the model's behavior by appropriate parameter selections
Internal validity	Run probabilistic models multiple times and check for consistency
Parameter variability-sensitivity analysis	Modify input parameters in a structured way and analyze the effects
Operational graphics	Display the values of performance measures for the system over time
Traces	Trace the behavior of specific objects in the model
Rationalism	Deduce the model logically

Table 2: Validation techniques in simulation-based research

To ensure the uniformity of the coding and to avoid ambiguity, the coding scheme was discussed intensively among all authors in a series of five workshops, totaling 11.5 hours, to reach a common understanding on each element of the given coding scheme. In the workshops, the authors discussed the underpinning criteria for each of the coding scheme's elements, relying on the provided arguments in the referenced studies upon which the analysis framework and, consequently, the coding scheme, is built. The latter resulted in the first version of a detailed guideline for coding. In a first step, we then conducted a pilot coding, in which two of the authors coded the same set of papers independently based on the initial coding guideline. Since the coding guideline mainly required judgment on the presence or absence of specific criteria in each paper, this was straightforward to do and reduced the potential for coders' subjective biases. In the second step, the authors discussed the few disagreements among coders in the pilot coding endeavor to ensure adequate inter-coder concordance. This brought about the development of the final coding guideline that was used by the first author to code all selected papers, including a re-coding of the pilot papers.

4 CURRENT STATE OF SIMULATION-BASED RESEARCH IN IS

Table 3 provides an overview of the selected 175 papers, showing during which time period these papers were published and which simulation techniques are employed. Most papers in this database are published in *Decision Support Systems* (65 papers), followed by *Information Systems Research* (33 papers) and the *Journal of Management Information Systems* (18 papers); see Table 6 in Appendix A1 for a detailed overview by outlet.

Simulation Technique	up to 1999	2000-2004	2005-2009	2010-2014	2015-2016	Total
Analytical	2	6	12	26	16	62
Agent-based	3	5	17	12	11	48
Stochastic Processes	1	2	6	16	10	35
System Dynamics	2	2	2	3	2	11
Artificial Neural Networks	0	2	3	3	2	10
Genetic Algorithms	0	1	3	2	3	9
Total	8	18	43	62	44	

Table 3: Overview of the use of simulation techniques over time

We focus our discussion on the most interesting insights resulting from a cross-element analysis of prior research, to illustrate how different modelling choices during simulation development are interrelated and

how they influence simulation use and validation. The structure follows the introduced elements of the analysis framework in the previous section: we first examine the use of theories as a foundation for simulation model development, then investigate several aspects related to the development of the simulation model itself, and finally describe how IS scholars have used and validated simulation models in their research. Table 4 summarizes our findings about simulation-based research in IS.

Real world to Simulation world	<i>Finding 1. Researchers increasingly rely on a wide range of system theories that can be easily translated into mathematical models, both from IS and from reference disciplines.</i>
	<i>Finding 2. Agent-based and analytical simulations generally employ multiple, complementary theories to describe different parts of the investigated socio-technical IS.</i>
	<i>Finding 3. Simulations based on artificial neural networks and genetic algorithms usually employ at most one system theory.</i>
Simulation world	<i>Finding 4. Agent-based and system dynamics simulations cover a comparatively large number of socio-technical system components.</i>
	<i>Finding 5. Artificial neural networks (and, to a lesser extent, analytical simulations, genetic algorithms, and stochastic processes) often cover only few socio-technical system components.</i>
Simulation world to Real world	<i>Finding 6. Agent-based (and, to a lesser extent, analytical and stochastic processes) simulations generally employ multiple, complementary validation techniques.</i>
	<i>Finding 7. System dynamics simulations often rely on adapting and combining existing models for validation.</i>
	<i>Finding 8. Simulations that use genetic algorithms often employ traces for validation.</i>
	<i>Finding 9. Artificial neural networks generally use historical data validation based on large datasets.</i>
	<i>Finding 10. Stochastic processes and analytical simulations are often used to prove or to critique.</i>
	<i>Finding 11. Genetic algorithms are often used to prescribe or to provide empirical guidance.</i>
	<i>Finding 12. Artificial neural networks are mostly used to predict or to prescribe.</i>
	<i>Finding 13. Agent-based simulations are commonly used to discover or to explain.</i>

Table 4: Summary of the literature review findings

Real world to simulation world: Our analysis of simulation-based IS research reveals that, similar to the trend in the IS discipline in general, the development of theory-based conceptual models in simulation-based studies is growing (Figure 2). Almost 80% of the analyzed papers that were published between 2015 and 2016 employ at least one theory in constructing their conceptual model. Compared to the 1990s and 2000s, the use of theories in simulation-based IS research has substantially increased in the 2010s, so that the more recent studies more frequently use theoretical lenses.

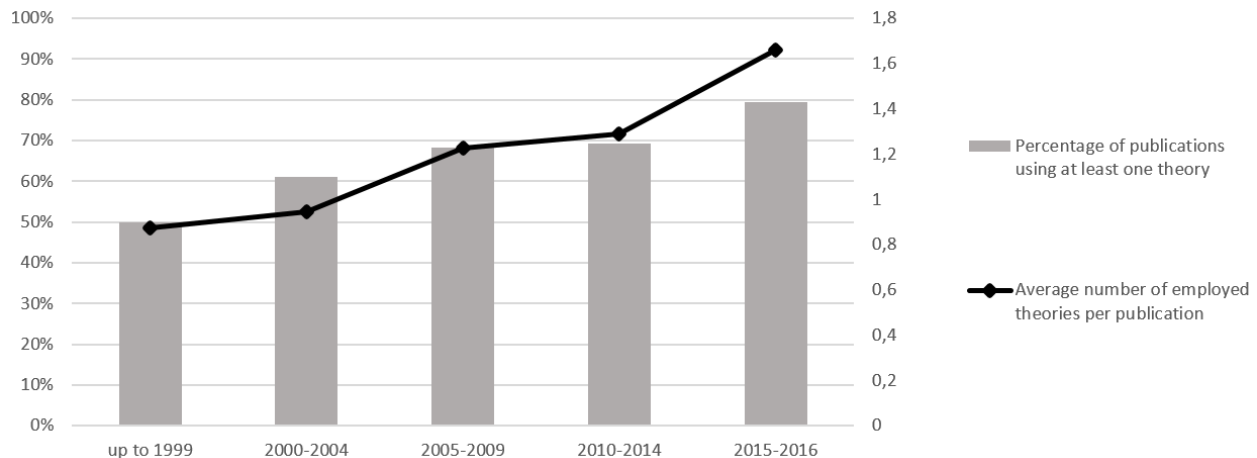


Figure 2: Use of theories in simulation-based IS research

A particularity of IS research is that, due to its multidisciplinary nature, a wide range of theories from reference disciplines, such as the organization, management, and computer sciences, are used in addition to native IS theories to guide both theory building and theory testing (Straub 2012; Lim et al. 2013). The same can be observed for simulation-based studies: the most frequently employed theories are both (i) commonly used theories in IS or native IS theories (e.g., *game theory*, the *technology acceptance model*, *competitive strategy*, or *portfolio theory*) as well as (ii) discipline-specific theories (e.g., *auction* and *queuing theories* in economics and computer science), which are easily applicable to the specific research questions of a given study. In effect, researchers exploit a wide range of theories that can be easily translated into mathematical models and that help researchers to systematically derive different scenarios subject to simulation (Finding 1 in Table 4).

Simulation Technique	Average number of theories used per publication	Percentage of publications using at least one theory
Artificial Neural Networks	0,70	60%
Stochastic Processes	0,89	54%
Genetic Algorithms	0,89	67%
System Dynamics	0,91	36%
Analytical	1,11	65%
Agent-based	1,83	83%

Table 5: Use of theories by simulation technique

Table 5 distinguishes different simulation techniques to investigate the average number of theories used in a single publication as well as the percentage of publications that use at least one theory. Combining these two data points allows us to make a number of interesting observations: first, we note that theories are mostly exploited in *agent-based* and *analytical* simulations (see Table 5). For these simulation techniques it is often the case that the complex socio-technical real-world system requires the use of multiple, complementary theories to describe different parts of the conceptual model (Finding 2 in Table 4). The same is true for *system dynamics* models in terms of using multiple theories, however most system dynamics models in our sample rely on adapting and combining existing models instead of referencing system theories. Nevertheless, *system dynamics* studies that employ system theories often use multiple theories.

These simulation techniques (*agent-based*, *analytical*, and *system dynamics*) can then be contrasted with *genetic algorithms* and *artificial neural networks*. We find that most studies rely on exactly one theory to justify input/output parameters (in the case of *artificial neural networks*) or parameters that define a candidate solution (in the case of *genetic algorithms*). Since the corresponding simulation models are essentially abstract, mathematical black boxes, they do not require multiple theories to translate complex socio-technical interactions into holistic conceptual models (Finding 3 in Table 4).

Simulation world: To analyze which components of a socio-technical system are covered by simulation models, we compare the percentages of publications that include a socio-technical system component for different simulation techniques (Figure 3). Unsurprisingly, the coverage of socio-technical system components often directly follows the nature of the employed simulation technique. For example, the autonomous agents in *agent-based models* usually match the socio-technical description of *actors*, and *system dynamics* models rely on a precisely defined *structure*.

Simulation Technique	Actors	Tasks	Structure	Environment	Technology	Average
Agent-based	●98%	●81%	●85%	●71%	○38%	●75%
System Dynamics	●73%	●73%	●100%	●82%	○27%	●71%
Analytical	●50%	●53%	○45%	○45%	○39%	●46%
Genetic Algorithms	○44%	●78%	○44%	○22%	○22%	●42%
Stochastic Processes	○46%	○46%	○40%	○31%	○31%	○39%
Artificial Neural Networks	○40%	○30%	○20%	○50%	○10%	○30%

Figure 3: Socio-technical components modeled by simulation technique

In Figure 3, we first can observe a larger coverage of socio-technical system components in *agent-based* and *system-dynamics* simulation models when compared to *stochastic processes* and *analytical* simulations, which explicitly model relations (Parunak et al. 1998; Carley 2001; Dooley 2002). To adequately represent a real-world problem, *agent-based* and *system dynamics* simulation models need to take a large number of factors into account, relating to different socio-technical components of the respective phenomenon and their interactions. Since the overall behavior of the simulation model emerges as a consequence of the modelled interactions, it is often a priori not clear, which factors are important and which factors may be ignored (Wu et al. 2015). Consequently, such models generally cover a comparatively large number of socio-technical system components (Finding 4 in Table 4).

On the other end of this spectrum are *artificial neural networks*, which do not require an explicit description of socio-technical interactions. Instead, the underlying models of this technique rely on abstract mathematical models that do not resemble any real-world socio-technical system components (Graupe 2013). Hence, we frequently find simulations of IS that employ *artificial neural networks* and only cover a limited number of socio-technical system components (Finding 5 in Table 4).

Simulation world to real world: We now analyze how simulations are validated and used in IS research. Figure 4 shows the percentages of papers that employ different validation techniques (not including *animation*, *degenerate tests*, and *turing tests*, which are not explicitly reported in any of the 175 analyzed papers) grouped by simulation technique. From this figure, we can see that *agent-based* simulations in particular rely on a wide range of different validation techniques to establish the credibility of the simulation model. The complex socio-technical nature of the investigated real-world IS phenomena augments the difficulties that researchers face in following the emergent process in which results are obtained in *agent-based* simulation experiments. Consequently, a single paper generally reports the use of multiple, complementary validation procedures for example, data driven techniques (e.g., *historical data validation*, *variability-sensitivity analysis*) as well as established knowledge (e.g., *comparison to other models*, *face validity*) and graphical support tools (e.g., *operational graphics*, *traces*). The same can be observed, to a lesser extent, for *analytical* and *stochastic processes* simulations that model complex socio-technical interactions over time (Finding 6 in Table 4).

Validation Technique	Agent-based	System Dynamics	Analytical	Stochastic Processes	Genetic Algorithms	Artificial Neural Networks
Comparison to Other Models	●90%	●91%	●53%	●43%	●22%	●30%
Internal Validity	●60%	○9%	●53%	●80%	●22%	●60%
Historical Data Validation	●42%	●36%	●48%	●31%	●78%	●100%
Variability-Sensitivity Analysis	●60%	○18%	●50%	●40%	●22%	●30%
Operational Graphics	●54%	●36%	●34%	●29%	●67%	●30%
Predictive Validation	●25%	○18%	●32%	○11%	●33%	●30%
Event Validity	●38%	●36%	●24%	○17%	○0%	○0%
Face Validity	●25%	●36%	○18%	○14%	○0%	○10%
Extreme Condition Tests	●31%	○18%	○10%	○9%	○11%	○10%
Traces	●23%	○0%	○5%	○9%	●56%	○0%
Rationalism	○4%	○9%	●23%	○6%	○0%	○10%

Figure 4: Employed validation techniques by simulation technique

Again, we note that most *system dynamics* models in our sample adapt or combine existing models in their construction, which is reflected in the high percentage for *comparison to other models* in Figure 4. The credibility of the extant models is then used to argue for modeling choices in a new simulation model (Finding 7 in Table 4). Regarding *genetic algorithms*, we find a surprisingly high percentage of publications that employ *traces*, i.e., track the behavior of specific objects during a simulation experiment. In these cases, researchers often follow the mutations and the development of a successful candidate solution and then use these observations to argue that the simulation model behaves as intended (Finding 8 in Table 4). Finally, simulations that rely on *artificial neural networks* commonly employ historical data (see Figure 4) to establish epistemic credibility in the simulation model. These papers essentially perform some type of cross-validation, where a dataset is split into training and validation parts. If the dataset is large enough and of sufficient quality, this is often enough to suggest that the simulation model behaves as intended, thus not requiring further validation efforts (Finding 9 in Table 4).

Figure 5 shows the percentage of papers employing a simulation technique for a specific use. More direct simulation techniques (*stochastic processes*, *analytical*) are often used to conduct *proofs*, in the sense that the simulation is employed to show that the modeled processes can produce a certain type of behavior (Harrison et al. 2007). Similarly, such simulation techniques can be employed to *critique* extant theoretical explanations of the observed phenomena, for example by describing more accurate or more parsimonious models. These simulation uses generally require a very precise mathematical description of the observed IS phenomena at the outset of the study (Finding 10 in Table 4).

Simulation Technique	Explanation	Discovery	Prediction	Prescription	Empirical Guidance	Critique	Proof
Agent-based	●77%	●75%	●48%	●33%	●42%	○6%	○8%
System Dynamics	●91%	●18%	●82%	●64%	●18%	○0%	○9%
Analytical	●39%	○3%	●35%	●58%	●42%	●23%	●52%
Stochastic Processes	●34%	○14%	●29%	●46%	●54%	●26%	●43%
Artificial Neural Networks	○10%	○10%	●80%	●60%	●80%	○0%	○10%
Genetic Algorithms	○0%	●22%	●33%	●67%	●89%	●33%	●33%

Figure 5: Simulation use by simulation technique

Similarly, genetic algorithms can be employed to demonstrate that certain adaptation strategies work (*proof*) or are better than previously suggested strategies (*critique*). Nevertheless, genetic algorithms are more dominantly used to provide *empirical guidance* by tracing the development of a successful candidate

solution to uncover the factors that determined its success. This technique is also frequently used to suggest more efficient designs of IS (*prescription*), generated through genetic mutations (Finding 11 in Table 4).

Artificial neural networks are mostly used to *predict* one or more output variables based on a set of multiple input variables. Along the same line, this simulation technique is frequently used to provide *empirical guidance* to testing multiple configurations of input variables, and use the results to identify the most important predictors for the studied phenomena. Another common use is the application of artificial neural networks in the design of specific IS (i.e., the simulation use *prescription*), which more efficiently manage certain tasks activities (Finding 12 in Table 4). For example, Kim and Street (2004) develop an artificial neural network that prescribes optimal customer targets for marketing.

In contrast, *agent-based* simulations are generally employed to *discover* unexpected consequences of the modeled interactions, and to *explain* the processes that produce the observed behavior. For example, Johnson et al. (2014) use simulation to study emergence mechanisms of social networks. The researchers start with several highly stylized descriptions of such mechanisms (e.g., social network structure emerges only through the rule of preferential attachment), which by themselves do not adequately represent the complex socio-technical interactions involved in the observed phenomena. Consequently, the researchers propose a blended, multi-theoretic model that better captures observed distributions in online networks (Finding 13 in Table 4).

5 DISCUSSION

We now reflect and discuss our findings from the literature review (see Table 4) to derive a set of four propositions that provide guidance for prospective simulation-based IS research. We again follow the structure of the preceding section (*real world to simulation world*, *simulation world*, and *simulation world to real world*), however, for our discussion we group the findings and relate them to the entire process of epistemic inferences. The goal is to point out the choices and consequences that researchers face, which we illustrate with specific exemplars of simulation-based IS research.

Simulation world to real world: Our review reveals that scholars increasingly refer to extant system theories in simulation-based research (see Figure 2) to support partially creative and intuitive modeling choices in the development of the conceptual model (Finding 1) (Winsberg 1999; Frank and Troitzsch 2005). Researchers generally opt for one of the following two choices of theories: (i) a rather comprehensive multi-theoretic foundation is employed, which is mainly used to develop agent-based or analytical simulations to discover or explain IS phenomena (Findings 2 and 13); or (ii) system theories are employed parsimoniously, which is particularly suitable for artificial neural networks and genetic algorithms (Findings 3, 11, and 12). Therefore, we propose:

Proposition 1. *The use of system theories guides the development of conceptual models in the investigation of IS phenomena through simulation-based research.*

Explication: *IS researchers often rely on a comprehensive multi-theoretic foundation. Multi-theoretic-based simulation models are particularly suitable for agent-based and complex analytical simulations as well as for discovery and explanation simulation uses. IS researchers may also employ system theories parsimoniously, using a specific theory to support the development of artificial neural networks and genetic algorithms for prediction and prescription.*

Simulations that study emergent phenomena in IS often employ a combination of theories to describe individual components of a system through complex socio-technical interactions (Woodard and Clemons 2014). Simulation-based studies purposefully combine complementary theories, i.e., a synergistic combination of theories that aims at a comprehensive analysis of the phenomenon of interest (Tiwana and Bush 2007). This is particularly important for agent-based, system dynamics, and complex analytical simulations, since in for these simulation techniques, the consequences of modeling complex socio-technical interactions are a priori unclear and minor deviations may result in fundamentally different model behavior

(Hedström and Ylikoski 2010; Grüne-Yanoff and Weirich 2010; Houy et al. 2012). The use of complementary theories helps scholars to justify modeling choices and also facilitates the interpretation of newly discovered insights in the simulation results (Morgan and Morrison 1999; Woodard and Clemons 2014; Bichler et al. 2016). An example of such research is the work of Nan (2011), where several theories of IT use (e.g., the structurational theory of technology and technology acceptance models) are combined with a complex adaptive systems approach to develop an agent-based simulation model (Nan 2011). Nan (2011) uses system theories to justify a wide range of modeling choices, ranging from models of mental activities of employees to organizational structures and environmental factors.

Conversely, Rivkin (2001) investigates the optimal level of strategic complexity in organizations by using a single, highly stylized fact. This stylized fact is based on case studies and prior theoretical work that links strategic complexity with a performance measure. This essentially follows the recommendation of Davis et al. (2007) to start with simple theory that addresses the phenomenon of interest and then directly translate this theory to a computational representation for a suitable simulation technique. Researchers may consult Bichler et al. (2016), who elaborate on the development of stylized facts based on existing system theories as well as Houy et al. (2015), who describe how stylized facts can be derived from extant literature and data.

Simulation world: Figure 3 shows that there are significant differences regarding the coverage of socio-technical IS components in different simulation techniques (Findings 4 and 5). Some simulation techniques, such as agent-based or system dynamics, rely on a description of interaction patterns on a local level to model system behavior (Bonabeau 2002). For these simulation techniques, we generally find a rather comprehensive coverage of socio-technical IS components in the simulation models. On the other hand, if a simulation technique relies on an abstract, formal mathematical model, a more specific coverage of socio-technical components may be better suited, as this facilitates the verification of the implemented simulation model and the subsequent validation and interpretation of simulation results (Adner et al. 2009). Thus, we propose:

Proposition 2. *Different simulation techniques require a different coverage of socio-technical IS components in the simulation model.*

Explication: *Agent-based and system dynamics techniques generally require a comprehensive coverage of socio-technical IS components in the simulation model. In contrast, artificial neural networks and, to a lesser extent, analytical simulations, genetic algorithms, and stochastic processes may employ simulation models that only cover a limited number of socio-technical IS components.*

This insight is in line with existing discourses on the creation of conceptual models: ontologically adequate conceptual models need to consider all potentially relevant aspects, including behaviors and co-evolutionary structures (McKelvey 2002; Houy et al. 2012), which is particularly true for simulation techniques that are commonly used to study emergent socio-technical phenomena. Nan (2011) to a wide extent covers socio-technical system components in the developed simulation model—comprising employees, tasks, information technology artefacts, organizational structures and environmental factors—, which is then used to experiment and to discover new explanations for the investigated phenomenon. Researchers may consult extant literature on socio-technical IS modeling (e.g., Lyytinen and Newman 2008; Bednar and Sadok 2015; Wu et al. 2015; Beese et al. 2015) and IS modeling in general (e.g., Houy et al. 2012; Frank et al. 2014; Frank 2014) to guide their simulation endeavors.

Real world to simulation world: The adequacy of a validation strategy is dependent on the specific simulation technique, the simulation use, the characteristics of the empirical target and the corresponding epistemic challenges (Boero and Squazzoni 2005). In general, it is recommended to combine internal verification with empirical validation in a circular process (Sargent 2005). This not only helps scholars to iteratively develop and test the simulation model, but also to validate obtained simulation results (Axtell et al. 1996; Edmonds and Hales 2003; Boero and Squazzoni 2005; Burton and Obel 2011). Literature on simulation model validation generally suggests the use of multiple validation techniques to build confidence in the connection of the simulation model to the underlying real-world system (Sargent 2005; Harrison et

al. 2007; Davis et al. 2007). This is in line with Figure 6, which shows that different simulation techniques rely on different validation strategies, which generally comprise multiple validation techniques (Findings 6, 7, 8, and 9). Hence, we propose

Proposition 3. *Investigating IS phenomena through simulation-based research, more specifically examining emergent behavior of IS phenomena, requires multiple validation techniques, depending on the employed simulation technique.*

Explication: *Artificial neural networks rely on the availability of large datasets and system dynamics models generally rely on adapting and combining existing models. Other simulation techniques, in particular agent-based simulations, rely on several complementary validation techniques that suit the chosen simulation technique as well as the available empirical data.*

Especially in the context of complex and dynamic socio-technical IS phenomena, the extent to which the simulation model accurately predicts and captures the essential behavior of the real-world system is often unclear. Some simulation approaches, such as artificial neural networks (see, e.g., Olson et al. 2012; Wang and Chuang 2016), offer a fairly straightforward way towards establishing epistemic credibility in the simulation model. To validate artificial neural networks, researchers may thus refer to established guidelines, for example consulting Arlot and Celisse (2010) on cross-validation procedures. In contrast, *epistemic opacity*, i.e. the inability of researchers to analytically trace how simulation results are obtained, is a notably difficult for indirect simulation techniques that aim at the discovery of novel phenomena and unexpected consequences (Harrison et al. 2007). Interpreting novel and unexpected simulation results requires a precise understanding of the dynamic processes in the simulation model (Burton and Obel 2011). These processes in the simulation model are often reflected in the action-formation and transformational mechanisms of the conceptual model, such as organizational and individual learning, decision-making, and imitation (Coleman 1994; Burton and Obel 2011). Issues in understanding emergent results of simulation experiments are related to existing discussions on computational models that go “beyond what is to explore possibilities and examine boundaries to what might be” (Burton and Obel 2011, p. 1197), essentially combining the issues arising from *equifinality* and the *opacity* of simulation experiments. If multiple assumptions may have led to the same outcome (*equifinality*), how can one argue that the observed outcome justifies the assumptions? Due to the inability of humans to follow all calculations in a simulation experiment in detail (*opacity*), such studies instead require a solid theoretical grounding and should employ multiple validation techniques to establish credibility in the simulation results (Winsberg 2001).

Finally, we move on to discussing simulation uses. In line with the work of Harrison et al. (2007) on simulation modeling in organizational research, we find that researchers in IS employ simulations for a variety of purposes, for which different simulation techniques may be suitable (Findings 10, 11, 12, and 13). Whereas stochastic processes and analytical simulations are frequently used to critique or prove, agent-based simulations generally aim at the discovery of new phenomena or at explaining previously made observations. Artificial neural networks and genetic algorithms are also frequently employed in the design of IS, fulfilling a more prescriptive purpose. We thus propose:

Proposition 4. *IS scholars can employ different simulation techniques, depending on the purpose of the simulation and the type of the targeted scientific contribution.*

Explication: *When aiming to prove or critique, IS researchers may consider simple stochastic processes or analytical simulation models. In contrast, agent-based models may be considered to discover new phenomena or explain the consequences of complex socio-technical interactions. For prescriptive research, such as design science, genetic algorithms and artificial neural networks can be considered.*

Reflecting this proposition in light of the earlier discussion on epistemic inferences in simulation-based research, we find that there is a fundamental difference between using simulation to predict empirical observations (step 5 in Figure 1) and to hypothesize (step 6 in Figure 1). This is reflected in the choice of simulation technique and consequently in the design of the simulation model. Using a simulation to prove, critique, or predict, is mostly complicated owing to epistemic issues related to the *predictive precision* of a simulation model: due to highly stylized abstractions and simplifications, the simulation model does not

necessarily show a precise correspondence with empirical data. Instead, simulation experiments rather aim to create statistical estimates of phenomena or observations that suggest the presence of certain conceptual relations, captured, for example through stylized facts (Küppers and Lenhard 2005). Simulations that aim to predict or prescribe, thus often focus on their utility, rather than their correspondence with reality, which are mainly conducted through artificial neural networks and genetic algorithms.

Conversely, agent-based simulations generally aim to discover new phenomena in complex socio-technical IS or to explore, refine, or contrast with scientific theory. Researchers usually test how different initial conditions and different processes lead to specific results in simulation experiments (Epstein 1999; Harrison et al. 2007; Davis et al. 2007; Weinhardt and Vancouver 2012; Poile and Safayeni 2016). Due to issues related to *equifinality* and *epistemic opacity*, these simulations need to be designed in a way that allow theoretical constructs to be traced back to these modeling choices. For example, Hua et al. (2011) propose an agent-based simulation model that investigates how combinations of operational decisions propagate through complex supply chain networks and finally lead to bankruptcy of organizations. For obtaining their results, they are required to conduct a series of highly structured simulation experiments that allow them to distinguish important decisions from unimportant or unrelated factors (Hua et al. 2011). Scholars interested in building such simulations may rely on a variety of extant knowledge, including, for example, research on visualization techniques (Zhuge 2006; Trier 2008) or on multilevel modeling and analysis (Bélanger et al. 2014; Frank 2014).

6 CONCLUSION

This study starts with the premise that although the presence of simulation-based research in IS discipline is relatively low, it has recently started to gain recognition within the IS community. This motivates us in considering the particularities of both IS, as a multidisciplinary research field, as well as simulation, as a third way of doing science compared to theoretical and empirical analyses. We first discuss the complex socio-technical nature of IS phenomena, which needs to be considered in simulation-based research, and the epistemic implications of using simulation-based research approaches. Building on these discussions, we derive an analysis framework to investigate the status quo of simulation-based research in IS. We finally synthesize the extracted findings on the current use of simulation in IS research to derive a set of four propositions, which aim not only to consolidate existing implicit knowledge about the use of simulation in IS, but also to guide prospective simulation-based IS research.

Regarding limitations, due to the breadth of the investigated subject, this research is of course selective in discussing related topics as well as limited in the depth of discussions. For each single simulation technique there exists a wealth of published knowledge and often we only scratch the surface in this paper. While we try to provide pointers to further information for readers, well-versed experts in a specific simulation technique will most likely have a better understanding of the corresponding intricacies than we present in this analysis.

Nevertheless, by studying a representative sample of simulation-based research in IS, we can use the tacit knowledge of experienced simulation researchers to facilitate future simulation-based IS studies for less experienced scholars. In terms of implications, the offered propositions provide the following guidance for prospective simulation-based IS research. *First*, the use of theoretical lenses (both the frequently used theories in IS and the native IS theories) is recommended to develop theory-informed simulation models. Depending on the employed simulation technique and the targeted scientific contribution, IS scholars can opt for exploiting multiple theoretical lenses (and make them a unified theoretical basis) or rely on single, simple theoretical groundings. *Second*, to account for the socio-technical nature of IS phenomena, some simulation techniques require a comprehensive coverage of socio-technical system components while other simulation techniques can more easily isolated and focus on specific components. *Third*, due to socio-technical and emergent character of IS phenomena, IS scholars should consciously and purposefully employ different validation techniques to ensure the reliability and validity of obtained simulation results. *Fourth*,

the choice of simulation techniques by IS scholars should be in line with the intended type of theorization in the given study. IS scholars should hence consciously opt for simulation techniques that are suitable for either a descriptive or prescriptive type of theorization (Gregor, 2006).

REFERENCES

- Adner R, Pólos L, Ryall M, Sorenson O (2009) The Case for Formal Theory. *Academy of Management Review* 34:201–208.
- Aminof B, Murano A, Rubin S, Zuleger F (2016) Automatic Verification of Multi-Agent Systems in Parameterised Grid-Environments. In: *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp 1190–1199
- Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. *Statist Surv* 4:40–79. doi: 10.1214/09-SS054
- Ashish AK, Aghav J (2013) Automated techniques and tools for program analysis. In: *Proceedings of the 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT)*.
- Axelrod R (1997) Advancing the Art of Simulation in the Social Sciences. In: *Simulating Social Phenomena*. Springer, Berlin, Heidelberg, pp 21–40
- Axtell R, Axelrod R, Epstein JM, Cohen MD (1996) Aligning simulation models: A case study and results. *Computational & Mathematical Organization Theory* 1:123–141.
- Becker J, Brocke J vom, Heddier M, Seidel S (2015) In Search of Information Systems (Grand) Challenges. *Business & Information Systems Engineering* 57:377–390.
- Bednar P, Sadok M (2015) A Socio-technical toolbox for business systems analysis and design. In: *1st International Workshop on Socio-Technical Perspective in IS Development (STPIS'15)*. CEUR, pp 29–31
- Beese J, Haki K, Aier S (2015) On the Conceptualization of Information Systems as Socio-Technical Phenomena in Simulation-Based Research.
- Bélanger F, Cefaratti M, Carte T, Markham S (2014) Multilevel Research in Information Systems: Concepts, Strategies, Problems, and Pitfalls.
- Bichler M, Frank U, Avison D, Malaurent J, Fettke P, Hovorka D, Krämer J, Schnurr D, Mueller B, Suhl L, Thalheim B (2016) Theories in Business and Information Systems Engineering. *Business & Information Systems Engineering* 58:291–319.
- Birdsey L, Szabo C (2014) An Architecture for Identifying Emergent Behavior in Multi-agent Systems. In: *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp 1455–1456
- Boero R, Squazzoni F (2005) Does Empirical Embeddedness Matter? Methodological Issues on Agent-Based Models for Analytical Social Science. *JASSS* 8.

- Bonabeau E (2002) Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences* 99:7280–7287.
- Bostrom RP, Heinen JS (1977) MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. *MIS Quarterly* 1:17–32.
- Bratley P, Fox BL, Schrage LE (2011) *A Guide to Simulation*. Springer Science & Business Media
- Burton RM, Obel B (2011) Computational Modeling for What-Is, What-Might-Be, and What-Should-Be Studies—And Triangulation. *Organization Science* 22:1195–1202.
- Carley KM (2001) Computational Approaches to Sociological Theorizing. In: Turner JH (ed) *Handbook of Sociological Theory*. Springer US, pp 69–83
- Coleman JS (1994) *Foundations of Social Theory*. Harvard University Press
- Curşeu PL (2006) Emergent states in virtual teams: a complex adaptive systems perspective. *J Inf technol* 21:249–261. doi: 10.1057/palgrave.jit.2000077
- Davis JP, Eisenhardt KM, Bingham CB (2007) Developing Theory Through Simulation Methods. *Academy of Management Review* 32:480–499.
- Doan TT, Yao Y, Alechina N, Logan B (2014) Verifying Heterogeneous Multi-agent Programs. In: *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp 149–156
- Dooley K (2002) Simulation research methods. In: Baum J (ed) *Companion to Organizations*. Blackwell, London, pp 829–848
- Dowling D (1999) Experimenting on Theories. *Science in Context* 12:261–273.
- D'Silva V, Kroening D, Weissenbacher G (2008) A Survey of Automated Techniques for Formal Software Verification. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 27:1165–1178.
- Edmonds B, Hales D (2003) Replication, Replication and Replication: Some Hard Lessons from Model Alignment. *JASSS* 6:4.
- Epstein JM (2008) Why Model? *JASSS* 11:4.
- Epstein LG (1999) A definition of uncertainty aversion. *Review of Economic Studies* 579–608.
- Frank U (2014) Multilevel Modeling. *Bus Inf Syst Eng* 6:319–337. doi: 10.1007/s12599-014-0350-4
- Frank U, Strecker S, Fettke P, Brocke J vom, Becker J, Sinz E (2014) The Research Field “Modeling Business Information Systems.” *Bus Inf Syst Eng* 6:39–43. doi: 10.1007/s12599-013-0301-5
- Frank U, Troitzsch KG (2005) Epistemological Perspectives on Simulation. *JASSS* 8:4.
- Frigg R, Hartmann S (2012) Models in science. In: *The Stanford Encyclopedia of Philosophy*, Spring 2012 edition. Stanford University, Stanford,

- Graupe D (2013) Principles of Artificial Neural Networks. World Scientific
- Gregoriades A, Sutcliffe A (2008) A socio-technical approach to business process simulation. *Decision Support Systems* 45:1017–1030.
- Grüne-Yanoff T, Weirich P (2010) The Philosophy and Epistemology of Simulation: A Review. *Simulation Gaming* 41:20–50.
- Harrison JR, Zhiang Lin, Carroll GR, Carley KM (2007) Simulation Modeling in Organizational and Management Research. *Academy of Management Review* 32:1229–1245.
- Hedström P, Ylikoski P (2010) Causal Mechanisms in the Social Sciences. *Annual Review of Sociology* 36:49–67.
- Houy C, Fettke P, Loos P (2015) Stylized Facts as an Instrument for Literature Review and Cumulative Information Systems Research.
- Houy C, Fettke P, Loos P (2012) Understanding Understandability of Conceptual Models – What Are We Actually Talking about? In: *Conceptual Modeling*. Springer, Berlin, Heidelberg, pp 64–77
- Hua Z, Sun Y, Xu X (2011) Operational causes of bankruptcy propagation in supply chain. *Decision Support Systems* 51:671–681. doi: 10.1016/j.dss.2011.03.007
- Humphreys P (1990) Computer Simulations. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1990:497–506.
- Jamroga W, Konikowska B, Penczek W (2016) Multi-Valued Verification of Strategic Ability. In: *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp 1180–1189
- Johnson SLF, SamerKudaravalli, Srinivas (2014) Emergence of Power Laws in Online Communities: The Role of Social Mechanisms and Preferential Attachment. *MIS Quarterly* 38:795-A13.
- Kim Y, Street WN (2004) An intelligent system for customer targeting: a data mining approach. *Decision Support Systems* 37:215–228. doi: 10.1016/S0167-9236(03)00008-3
- Kochanowicz J, Ah-Hwee T, Thalmann D (2013) Enabling Generative, Emergent Artificial Culture. In: *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp 789–792
- Kulik BW, Baker T (2008) Putting the organization back into computational organization theory: a complex Perrowian model of organizational action. *Computational & Mathematical Organization Theory* 14:84–119.
- Küppers G, Lenhard J (2005) Validation of Simulation: Patterns in the Social and Natural Sciences. *JASSS* 8:4.
- Law AM (2008) How to Build Valid and Credible Simulation Models. In: *Proceedings of the 40th Conference on Winter Simulation*. Winter Simulation Conference, Miami, Florida, pp 39–47

- Lee AS (2010) Retrospect and prospect: information systems research in the last and next 25 years. *Journal of Information Technology* 25:336–348.
- Lee J-S, Filatova T, Ligmann-Zielinska A, Hassani-Mahmooei B, Stonedahl F, Lorscheid I, Voinov A, Polhill JG, Sun Z, Parker DC (2015) The Complexities of Agent-Based Modeling Output Analysis. *JASSS* 18:4.
- Lim S, Saldanha T, Malladi S, Melville N (2013) Theories Used in Information Systems Research: Insights from Complex Network Analysis. *JITTA* 14:2.
- Loos P, Mettler T, Winter R, Goeken M, Frank U, Winter A (2013) Methodological Pluralism in Business and Information Systems Engineering? *Bus Inf Syst Eng* 5:453–460.
- Lowry PB, Moody GD, Gaskin J, Galletta DF, Humphreys S, Barlow JB, Wilson D (2013) Evaluating Journal Quality and the Association for Information Systems (AIS) Senior Scholars' Journal Basket Via Bibliometric Measures: Do Expert Journal Assessments Add Value? *Social Science Research Network*, Rochester, NY
- Luna-Reyes LF, Zhang J, Ramón Gil-García J, Cresswell AM (2005) Information systems development as emergent socio-technical change: a practice approach. *Eur J Inf Syst* 14:93–105. doi: 10.1057/palgrave.ejis.3000524
- Lyytinen K, Newman M (2008) Explaining information systems change: a punctuated socio-technical change model. *Eur J Inf Syst* 17:589–613.
- Macal CM, North MJ (2009) Agent-based Modeling and Simulation. In: *Winter Simulation Conference. Winter Simulation Conference*, Austin, Texas, pp 86–98
- Manzo G (2007) Variables, Mechanisms, and Simulations : Can the Three Methods Be Synthesized ? *Revue française de sociologie* 48:35–71.
- McKelvey B (2002) Model-centered organization science epistemology. In: *Companion to organizations*. Oxford University Press, New York, pp 752–780.
- McLeod L, Doolin B (2012) Information systems development as situated socio-technical change: a process approach. *Eur J Inf Syst* 21:176–191. doi: 10.1057/ejis.2011.43
- Montali M, Calvanese D, De Giacomo G (2014) Verification of Data-aware Commitment-based Multiagent System. In: *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp 157–164
- Morgan MS, Morrison M (1999) *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge University Press.
- Nan N (2011) Capturing Bottom-up Information Technology Use Processes: A Complex Adaptive Systems Model. *MIS Quarterly* 35:505-A7.
- Olson DL, Delen D, Meng Y (2012) Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems* 52:464–473. doi: 10.1016/j.dss.2011.10.007
- Orlikowski WJ (1992) The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science* 3:398–427. doi: 10.1287/orsc.3.3.398

- Orlikowski WJ, Iacono CS (2001) Research Commentary: Desperately Seeking the “IT” in IT Research—A Call to Theorizing the IT Artifact. *Information Systems Research* 12:121–134. doi: 10.1287/isre.12.2.121.9700
- Parunak HVD, Savit R, Riolo RL (1998) Agent-Based Modeling vs. Equation-Based Modeling: A Case Study and Users’ Guide. In: Sichman JS, Conte R, Gilbert N (eds) *Multi-Agent Systems and Agent-Based Simulation*. Springer Berlin Heidelberg, pp 10–25.
- Pias C (2011) On the Epistemology of Computer Simulation. *Zeitschrift für Medien- und Kulturforschung* 2011:29–54.
- Poile C, Safayeni F (2016) Using Computational Modeling for Building Theory: A Double Edged Sword. *JASSS* 19:3.
- Prat N, Comyn-Wattiau I, Akoka J (2015) A Taxonomy of Evaluation Methods for Information Systems Artifacts. *Journal of Management Information Systems* 32:229–267. doi: 10.1080/07421222.2015.1099390
- Ripley BD (2009) *Stochastic Simulation*. John Wiley & Sons
- Rivkin JW (2001) Reproducing Knowledge: Replication Without Imitation at Moderate Complexity. *Organization Science* 12:274–293. doi: 10.1287/orsc.12.3.274.10106
- Roberts N, Deal R, Shaffer W, Garet M, Anderson D (1994) *Introduction to Computer Simulation; A System Dynamics Approach*. Productivity Press Inc.
- Rohrlich F (1990) Computer Simulation in the Physical Sciences. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1990:507–518.
- Sargent RG (2005) Verification and Validation of Simulation Models. In: *Proceedings of the 37th Conference on Winter Simulation*. Winter Simulation Conference, Orlando, Florida, pp 130–143.
- Sidorova A, Evangelopoulos N, Valacich J, Ramakrishnan T (2008) Uncovering the Intellectual Core of the Information Systems Discipline. *MIS Quarterly* 32:467–482.
- Spagnoletti P, Za S, Winter R (2013) Exploring foundations for using simulations in IS research. In: *Proceedings of the 34th International Conference on Information Systems*. Milan, Italy.
- Sterman JD (2002) Systems dynamics modeling: tools for learning in a complex world. *IEEE Engineering Management Review* 30:42–42. doi: 10.1109/EMR.2002.1022404
- Straub D (2012) Editor’s Comments: Does MIS Have Native Theories? *Management Information Systems Quarterly* 36:iii–xii.
- Tiwana A, Bush AA (2007) A Comparison of Transaction Cost, Agency, and Knowledge-Based Predictors of IT Outsourcing Decisions: A U.S.-Japan Cross-Cultural Field Study. *JMIS* 24:259–300.
- Trier M (2008) Research Note—Towards Dynamic Visualization for Understanding Evolution of Digital Communication Networks. *Information Systems Research* 19:335–350.
- vom Brocke J, Simons A, Riemer K, Niehaves B, Plattfaut R, Cleven A (2015) Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research.

- Wang C-H, Chuang J-J (2016) Integrating decision tree with back propagation network to conduct business diagnosis and performance simulation for solar companies. *Decision Support Systems* 81:12–19. doi: 10.1016/j.dss.2015.10.004
- Webster J, Watson RT (2002) Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly* 26:xiii–xxiii.
- Weinhardt JM, Vancouver JB (2012) Computational models and organizational psychology: Opportunities abound. *Organizational Psychology Review* 2:267–292.
- Whitley D (1994) A genetic algorithm tutorial. *Stat Comput* 4:65–85. doi: 10.1007/BF00175354
- Winsberg E (1999) Sanctioning Models: The Epistemology of Simulation. *Science in Context* 12:275–292.
- Winsberg E (2003) Simulated Experiments: Methodology for a Virtual World. *Philosophy of Science* 70:105–125.
- Winsberg E (2001) Simulations, Models, and Theories: Complex Physical Systems and Their Representations. *Philosophy of Science* 68, pp.442–454.
- Wöhner T, Köhler S, Peters R (2015) Managed Wikis - A New Approach for Web 2.0. *Business & Information Systems Engineering* 57:155–166.
- Woodard CJ, Clemons E (2014) Modeling the Evolution of Generativity and the Emergence of Digital Ecosystems. In: *Proceedings of the 35th International Conference on Information Systems*.
- Wu PP-Y, Fookes C, Pitchforth J, Mengersen K (2015) A framework for model integration and holistic modelling of socio-technical systems. *Decision Support Systems* 71:14–27.
- Xiao S, Dong M (2015) Hidden semi-Markov model-based reputation management system for online to offline (O2O) e-commerce markets. *Decision Support Systems* 77:87–99. doi: 10.1016/j.dss.2015.05.013
- Zhang M, Gable G (2014) Rethinking the Value of Simulation Methods in the Information Systems Research Field: A Call for Reconstructing Contribution for a Broader Audience. In: *Proceedings of the 35th International Conference on Information Systems*.
- Zhuge H (2006) Knowledge flow network planning and simulation. *Decision Support Systems* 42:571–592. doi: 10.1016/j.dss.2005.03.007

APPENDIX A1: DESCRIPTION OF LITERATURE DATABASE

Outlet	Initial search	After first step	Final database
Business and Information Systems Engineering	12	2	1
Decision Support Systems	281	86	65
European Conference on Information Systems	37	13	8
Electronic Commerce Research and Applications	33	5	4
European Journal of Information Systems	6	6	3
International Conference on Information Systems	36	13	6
Information Systems Frontiers	53	12	10
Information Systems Journal	4	4	1
Information Systems Management	8		
Information Systems Research	39	39	33
Information Technology and Management	21	5	4
International Journal of Electronic Commerce	14	2	1
Journal of Computer Information Systems	25	1	0
Journal of Database Management	7		
Journal of Global Information Systems	1		
Journal of Information Technology	5	5	2
Journal of Management Information Systems	26	26	18
Journal of Organizational Computing and Electronic Commerce	17		
Journal of Strategic Information Systems	5	5	2
Journal of the Association for Information Systems	5	5	3
Management Information Systems Quarterly	26	26	14
Management Information Systems Quarterly Executive	1		
Wirtschaftsinformatik	35		
Total	697	255	175

Table 6: Overview of selected papers by outlet