Scandinavian Journal of Information Systems

Volume 26 | Issue 1 Article 1

6-30-2014

How Critical Realism Clarifies Validity Issues in Information Systems Theory-Testing Research

Stephen P. Smith
Monash University and The University of Melbourne, stephen.smith@monash.edu

Robert B. Johnston
The University of Sydney and Monash University, Robert.Johnston@ucd.ie

How Critical Realism Clarifies Validity Issues in Information Systems Theory-Testing Research

Cover Page Footnote
We thank Margunn Aanestad and three anonymous reviewers for their very helpful feedback and also their patience while we implemented those suggestions.

How Critical Realism Clarifies Validity Issues in Information Systems Theory-Testing Research

Stephen P. Smith
Monash University and The University of Melbourne, Australia
stephen.smith@monash.edu

Robert B. Johnston
The University of Sydney and Monash University, Australia
robert.johnston@ucd.ie

Abstract. Validity is an enduring theme in the information systems (IS) domain. However, because much of that discussion draws, at least nominally, on an Empiricist orientation, formulating issues of validity in a way that is pertinent to a practice-oriented discipline has been difficult. This is particularly true for external validity, which refers to the applicability, or relevance, of research to phenomena in the broader environment of practice, but also for measurement, ecological, and internal validity, all of which are needed to establish rigor in human-centred practical research. We argue that critical realism presents a number of insights that profoundly clarify a discussion of validity in practice-oriented theory-testing research. We reformulate the notions of theory testing and research validity from the critical realist perspective and illustrate new insights gained using a discussion of a controlled experiment.

Keywords: Critical realism, validity, hypothesis-testing, theory.

1 Introduction

Ensuring that research is valid is an enduring theme in the information systems (IS) domain, as evidenced by the large number of articles in our premier journals that focus only on that topic (King and He 2005; Klein and Myers 1999; Lee and Baskerville 2003; O'Leary-Kelly and Vokurka 1998), and the amount of column space in more conventional theory-testing articles that purports to demonstrate that the work is valid in some way. A significant amount of this research uses a quantitative hypothesis-testing approach with a heavily positivist slant. That style

Accepting editor: Margunn Aanestad

of research typically discusses validity at some length, arguing that the measures, the sampling strategy, and the data analysis are all valid in some way, and that the results apply beyond the sample of observations reported.

Discussions are nevertheless challenging; we argue that the challenge stems from certain inherent contradictions and inconsistencies associated with the traditional approach to assessing validity issues in theory-testing research. For information systems, which is principally a practice-based research domain, there is the further problem of ensuring that research provides insight into problems faced by the community of practitioners. This is particularly true for external validity, which refers to the applicability, or relevance, of research conducted in contrived or bounded research situations to phenomena in the broader environment of practice, but it is also (to varying extents) true of measurement, ecological, and internal validity, all of which seek to establish the rigor of human-centered practical research.

Within each study, researchers have to be aware that human agency must be accounted for in any explanation of the operation of an information system because system processes and outcomes will be influenced by the goals of the people who build and operate that system. In other words, research into systems-related issues is complicated because the operation of the system, unlike a natural force, is not independent of the conceptions of the people who use and are governed by the system. This added complexity, we will argue, is most problematic for researchers who adopt a rigorous empiricist stance when describing how validity issues are satisfied. The impoverished nature of that philosophy has resulted in IS researchers and writers making ad hoc, modifications to the empiricism, arguably because the practice orientation of IS seems to require a more realist approach to sciences and social sciences, but these are only partly convincing because they are generally not derived from a consistent theoretical base.

In this paper we aim to show that new clarity can be brought to the formulation of the traditional validity issues of theory-testing research in IS by explaining the fundamental principles of each issue in the language of the critical realist philosophy of Bhaskar (1978; 1993; 1998a), Sayer (1992), Collier (2011), Archer (2013), and others. Critical realism is increasingly discussed in IS as a grounding philosophy for research methodology that both breaks with hard empiricist approaches and accepts the epistemological criticism of the interpretivists, while still retaining a realist ontology suited to scientific research (e.g., Tsang 2013; Volkoff and Strong 2013; Wynn and Williams 2012). Having these properties means that this research philosophy should provide a robust basis for clarifying methodological issues in theory-testing research. Our focus on validity issues in theory-testing research adds to the growing body of literature on the research process from a critical realist perspective, including how critical realism resolves inconsistencies between the ontological assumptions and research practices of much IS research (Smith 2006), how it informs the research question or focus of IS studies (Dobson 2002), and the process of developing and refining theory, particularly in system evaluation studies (Carlsson 2003; Carlsson et al. 2010), how it can be used to improve the rigor of case study research (Wynn and Williams 2012) and why diversity in data gathering and analysis methods is desirable (Mingers 2004a).

We understand that most current IS researchers would probably reject the strong form of empiricism that we describe by way of example throughout this paper. Actual research practice and pedagogy already incorporate many of the insights that we find follow from a critical realist perspective (Dobson 2002; Mingers 2000; Smith 2006). Our point is that these extensions lack an adequate philosophical foundation so long as empiricism continues to be the framework used for their exposition. This is especially true for questions of validity. Our thesis is that critical realism provides such a precise alternative foundation.

Our argument proceeds as follows. In the next section we outline briefly the epistemology of empiricism. The goal here is not to present a detailed critique of that philosophy; rather, we wish to demonstrate the poverty of this position for discussion of research validity. We then outline critical realism to identify how this position allows the problems of theory testing and research validity to be reformulated with new clarity, with that approach systematized in the following section. Building on these principles, we describe a hypothetical controlled experiment that illustrates the theoretical analysis and explicates the steps that might be taken to ensure and justify validity according to the new critical realist formulation. We conclude with a summary of the significance of our arguments and their broader implications.

2 The standard empiricist view of theory testing

Enquiry in the natural and social sciences is influenced heavily by the empiricist world view, which holds that all valid knowledge is empirical and so ultimately derived from sensory experience (Hooker 1975). Hume's scepticism, to the effect that no knowledge can be established conclusively by reason alone, has been particularly influential in the development of scientific methods (Fogelin 1993), and is the foundation for the hypothetico-deductive approach that is so prevalent in the natural and social sciences (Rosenberg 1993).

Logical positivism is an extreme extension of empiricism that rejects all claims of knowledge except those derived empirically. On this view, because the only meaningful form of knowledge is empirical (and ultimately derived from the sensory experience that informs direct knowledge), the ultimate aim of science is to maximize empirical knowledge (Hooker 1975). In keeping with Hume's view that what we think of as a causal relationship is merely a regularity in perceptions (a constant conjunction of events), a statement that one event causes another is equivalent to saying that one event is always followed by another. Under this highly influential world view, causation cannot be proved, so the proper role of theory is for predicting patterns based on objective observations.

Theory testing, from an empiricist standpoint, finds its ultimate expression in Popper's falsificationism (Popper 1959), which holds that because we can never prove a theory definitively, a scientific hypothesis must be evaluated in terms of whether it is actually false. Theory-testing, on this view, means assessing observational evidence to determine whether (1) observations are consistent with an a priori hypothesis, and (2) this correspondence is systematic and unable to be explained plausibly by chance factors alone (the null hypothesis). This model of theory-testing is illustrated in figure 1, which shows the relationship between the empirical and theoretical domains under that paradigm. Note that the diagram is a more nuanced view of the research process than is possible using Hume's modern empiricism, and is (arguably) a reflection of the neo-empiricist view of research (see Greenwood 1990).

The shortcomings of this philosophy, particularly that it tolerates only an impoverished view of the world, have been dealt with at length (Klein and Lyytinen 1985; Kolokowski 1972; Mingers 2004a; Smith 2006; Williams 1975), and so are not the focus of our discussion. For

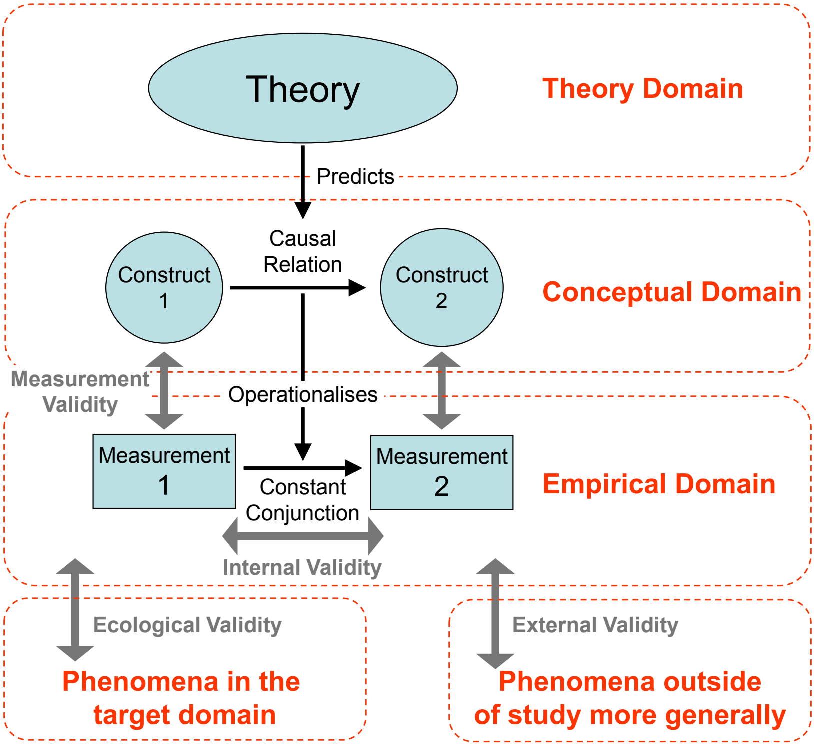

Figure 1. Research process and validity according to the Empiricist research tradition (after Neuman 2005)

the purposes of this discussion, however, two points should be noted regarding figure 1. Firstly, on a strict reading of empiricism, only empirical and theory categories enter the empiricism paradigm. Limiting the objects of enquiry in this way implies that the objective of science is merely to gather and predict data, and not to explain at any deep level the mechanism that enabled those experiences (because nothing beyond experience can be analysed validly). Secondly, limiting the research focus to a comparison of experience with theory means that measurement validity, internal validity, ecological validity, and external validity must all be primarily concerned with empirical issues.

Measurement validity, for the empiricist, is primarily concerned with attributes of the data, especially statistical bias, coherence, and differentiation. Under that view, measurement validity is established by demonstrating, for example, that parts of the measure are consistently related to other measures that are conceptually similar, and distinct from measures that are conceptually dissimilar.

Internal validity is concerned with showing that there is a constant conjunction between particular observations and events. For an experiment, this means demonstrating that any difference in the events measured is systematically related to changes in the variable being studied, and that the constructed environment forms a closed system by virtue of the experimental controls. To the extent that closure can be demonstrated, differences in the pattern of observed event sequences can be attributed to the research intervention.

Ecological validity is an acknowledgement that human activities are situated in, and influenced by, particular contexts. From the empiricist's perspective, because research is focused on events and measures, it is important that the environment of the study (including the look and feel of stimuli) corresponds to a natural setting because results should then reflect the outcomes that should occur in that natural setting.

External validity is whether the researcher can "infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings, and times"(Cook and Campbell 1979). Applying a pattern or rule to settings beyond those that have been observed is problematic from an empiricist perspective, however, because the reasoning process required (applying observations to unobserved phenomena) violates fundamental principles of empiricism. Specifically, empiricism requires the assumptions of any model to have been tested empirically before being included in that model. However constructing any argument about the validity of assumptions outside of the specific data set examined (or other known data sets) ultimately requires the use of inductive reasoning, a form of analysis that empiricism regards as invalid because the validity of inductively derived assertions cannot be tested directly through observation.

The limitations of the empirically-centered conception of validity are highlighted in the discussion of this issue in some positivist-oriented IS literature. For example, Lee and Baskerville's (2003) lengthy discussion about non-statistical forms of generalizability essentially proposes pragmatic workarounds to some of the serious validity constraints that a strict empiricist approach imposes on a practice-focused discipline. Also see various critiques on validity-related issues in research (Boudreau et al. 2001; Klein and Lyytinen 1985).

From these discussions and the standard definitions of validity terms presented, it is clear that:

1. *Measurement validity* is problematic in IS research because many of the things we would like to measure, such as system usefulness or system success, have socially constructed or experiential aspects, and there is no direct way to assess the correspondence between a measure and a behaviour. Moreover, in circumstances where it is possible to observe an outcome, as with actual use of a technology, there is often a close relationship between data, measures, and constructs. In this situation, there is a tendency to conflate a trait with the test for that trait (assuming that a variable used in analysis exists in the world), resulting in its theoretical properties being defined in terms of data values (Bunge 2006; Mingers 2003). An example from the IS domain is *usefulness.* There is no obvious way to observe usefulness, particularly each individual's socially constructed understanding, so the design of the measure is critical for gaining access to the underlying reality. However, one cannot simply use a Likert-type scale with some concepts thought to be related to usefulness, run factor analysis, and keep the items that display high levels of convergent and discriminant validity. There is no guarantee, using such an approach, that the items actually measure the concept of interest, even if they are intercorrelated and empirically distinct from other measures. If we instead base our assessment on observation of what

someone does when using a system, we then run into the problem of separating a test for usefulness from our observations of system outcomes.

2. *Internal validity* is an apparently straightforward concept in experimental studies, but relies on demonstrating that the study environment is a closed system with respect to the manipulations. The problematic aspect of internal validity is not that it is difficult to achieve, but rather that measurement and effect tend to be conflated, leading to a lack of differentiation between a stimulus (or manipulation) and the underlying reality represented (Bunge 2006).

3. *Ecological validity* is problematic to the extent that research results are regarded as unreliable unless the environment of the study mimics a typical setting for the activities studied. In other words, a specific naturalistic setting is needed in an experiment to demonstrate that a particular regularity can occur in the so called *real world* setting modelled. A more subtle problem is that, even when a setting is made naturalistic, the precise operational definitions demanded by a philosophical emphasis on empirics can so deplete a setting (the instance of practice modelled) of its richness and texture that it all but disappears (Madill et al. 2000).

4. The epistemological status of *external validity* is problematic for researchers in a practice-focused discipline, such as IS. Because our focus is on informing practice, we typically collect data about events in order to build theories and make recommendations that are potentially of assistance to practitioners. A limited set of environments is available to us in any given project, of course, but lessons drawn from the data are rarely restricted to just what was observed. Indeed, an important test of the success of any theory is whether it is useful for solving related (but non-identical) technology-related problems or assisting with managing resources in other (unobserved) environments. Empiricism falls short, because it does not really allow for the type of reasoning required to make the generalizations about the applicability of assumptions to unobserved objects that is required for a researcher to claim external validity (Lucas 2003).

Studies of the actual methods used by researchers in the information systems, management, and social psychology disciplines indicate widespread departures from strict empiricist requirements, creating an implicit tension between ideal and actual practice, whereby researchers claiming to follow an empiricist tradition are unable to explain some of their research practices in strict empiricist terms (see Bealer 1992; Maddy 1992; Smith 2006). In information systems research in particular, these differences indicate that critical realism may be a more appropriate formulation of what information systems researchers actually believe and practice (Mingers 2004b; Smith 2006). The realist position we expound in the next section illuminates this issue, and provides a set of conceptual tools that help address many of the problems that IS researchers have trouble dealing clearly with from the espoused empiricist position. In addition, by reformulating validity using a sophisticated realist philosophy we are able to demonstrate that validity is not primarily an empirical issue. Note that our description of the basic principles of critical realism is limited to the terminology that is strictly relevant to this work. For a more general primer on the critical realist philosophy as it relates to information systems see work by Mingers (Mingers 2004b; Mingers et al. 2013).

3 Brief outline of critical realism

Bhaskar's philosophy of critical realism (Bhaskar 1978; 1998a), as its name suggests, is a realist philosophy, which is to say that it claims that a world outside and independent of our conscious perception exists, and that only some aspects of this world are objectively knowable via our senses. Our senses are not completely reliable, of course; for example, we can be fooled by illusions. Nevertheless, because reality is independent of our senses, even when we misperceive an event, the occurrence and properties of that event are independent of our perception and understanding, and the cause of the event operated even if we were not aware of its operation.

This distinction between what happens and what we perceive, and between an event and the underlying (but possibly unobservable) mechanism that caused that event, are the key aspects of critical realism that we will explore here. In the language of critical realism, they form three strata of reality: the real domain, the actual domain, and the empirical domain. The actual domain is the easiest to describe. This is the domain of events: someone forms an intention to sell shirts via an Internet-based store, a consumer visits that store, and so on.

The empirical domain is what people can experience. Things in the actual world, such as events cannot be perceived, but events leave empirical traces including perceptions. That is, the sensing of an event is not the same as the event, which may occur regardless of whether we can sense it. In addition, the subjective and perspectival nature of perception means that experiences will vary from one person to another and across settings.

According to critical realism, behind events are structures and generative mechanisms that have enduring properties. In non-technical terms, a generative mechanism is the causal power that gives rise to something, or the reason that something *is* (Bhaskar 1993). The hierarchy of ontological categories, then, is that enduring generative mechanisms and structures are part of the overarching category of the real. These mechanisms and structures instantiate actual events (and possible non-occurrences of events), which leave empirical traces that can be observed or otherwise experienced. Therefore, mechanisms, events, and experiences are all real. Events and experiences are also actual (because they are particular occurring instantiations of the generative mechanisms), but because experiences are obtained via empirical traces of actual things, experiences are also empirical. The nested relationship between these three domains of the real, representing the stratified ontology of critical realism, is illustrated in figure 2, derived from Mingers (2004b, p. 94).

From the critical realist perspective, understanding the real domain is the proper role of science. To develop theory, from this perspective, is to explain *why* an event occurs, but from a transcendental perspective. That is, the focus is not usually on the specific event observed, but on what that event tells us about enduring underlying causal relationships (generative mechanisms) that lie beyond common experience (the empirical domain). Scientific investigation, on this view, involves manipulation of the environment to trigger or manipulate the operation of a generative mechanism in order to produce particular actualised outcomes. For example, a re-

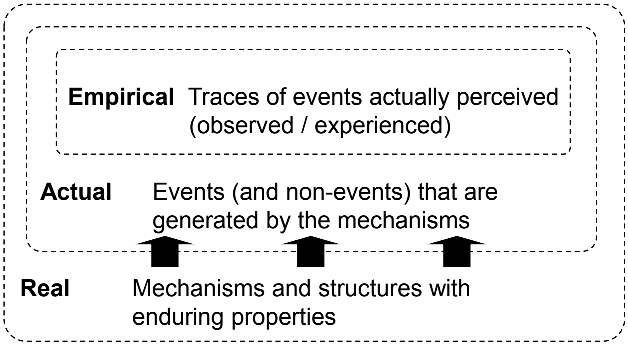

Figure 2. The stratified ontology of critical realism, after Mingers (2004a, p. 94)

searcher could manipulate the environment surrounding a particular class of decision process to answer the questions why would someone want to buy a shirt online, and what makes a shopper select one online vendor in preference to another? Further, because critical realism distinguishes between the cause, the event, and data about that event, any causal explanation must explain patterns of events independently of any particular event or data about that event. In the language of critical realism, these causal forces are known as *generative mechanisms*.

To explain the nature and workings of a generative mechanism, critical realism asks the researcher to use retroductive reasoning. Retroduction, also sometimes referred to as abduction, is a mode of reasoning in which the cause of an event is inferred, based on the question 'what must the world be like for the event to be possible?' and the converse, 'what must the world be like for the event to not be possible?' The explanation may in fact be incorrect, but it is nevertheless a useful starting point for more analysis in which new data is gathered and the explanation retested. The objective of this type of reasoning is to determine an explanation for observations, particularly with a view to identifying the circumstances without which those observations cannot exist, based on the assumption that nature's uniformity is not the result of accidental regularities, but rather a consequence of the existence and operation of underlying mechanisms and structures (Sayer 1992, p. 158). Bhaskar (1978) argues that retroduction describes the iterative approach actually used to develop most theories in the sciences and social sciences because it involves looking deeply into (or behind) observed patterns or regularities to discover what produces them.

With this background we can now see how this philosophical orientation applies to objects of knowledge in the natural sciences, where an event and empirical traces of that event can be explained and predicted by a generative mechanism. Knowledge, under this philosophical paradigm, is accumulated by the iterative process of identifying systematically the entities responsible for an event, proposing a generative mechanism (via retroductive reasoning), testing that theory, and repeating all or part of that process to address shortcomings (Bhaskar 1978).

These principles also apply to the social sciences where the objects of enquiry are socially constructed entities, such as language, society, and social capital. The transitive nature of the structures examined there (compared to the intransitive structures examined in the natural sciences) creates a number of ontological and epistemological issues. For example, ontologically, social structures do not exist independently of their effects, are localized in time and space, and only hold for particular contexts; and epistemologically, social science is self-referential in the sense that like the phenomena studied, it is a social practice (Mingers 2004a). However, these structures nevertheless exist independently of the researcher and the research activity, and are causal and therefore real. The job of the social scientist therefore, is to explain the nature and operation of these social structures, objects, and generative mechanisms.

In a practice-based discipline, such as IS, explanation of phenomena relevant to members of the community is similarly required to develop and advance the shared and specialised body of knowledge that informs activities. This means that an explanation of *how* a given variable could have influenced a practice-relevant event is needed (e.g., *why* a lack of top management support may lead to implementation problems), not just a description of empirical relationships, particularly given that empirical relationships may only apply in a limited range of circumstances (and in some cases may merely be statistical anomalies). In fact, practice-based disciplines are arguably at least as dependent on the type of deep explanations demanded by critical realism as scientists and social scientists because practice activities are informed by a body of abstract knowledge that (at least in part) is derived from enquiry in the social and natural sciences.

4 Implications of critical realism for theory testing in IS

The three-fold stratified ontology just presented, differs substantially from both positivism and interpretivism, the dominant espoused research paradigms in the information systems discipline (Smith 2006), but also offers profound insight into the nature of the research activities we, as a practice-oriented discipline, undertake. We will now describe how the principles of critical realism assist our understanding of validity issues in research.

4.1 Principles that assist our understanding of validity issues

We identify four principles of critical realism, summarised in table 1, that help to clarify what constitutes valid research. Firstly, *critical realism distinguishes the generative mechanism from the operation of that mechanism in specific circumstances.* In other words, events are caused by the operation of mechanisms, and are no less real than the mechanisms that cause them, but they are ontologically distinct. The types of events that we investigate in the information systems domain tend to be social phenomena (e.g., IS success and failure). This means that a given outcome, say a decision to abandon an IT infrastructure project, is distinct from the underlying mechanism, which we might call the *project failure mechanism*, and which can result in both successful and unsuccessful outcomes depending on the specific circumstances of a project.

Secondly, *critical realism distinguishes between theory and the generative mechanism that the theory describes.* This distinction between the theory and the generative mechanism to which it refers is the key to understanding how the implications of critical realism for theory testing and validity differ from those of empiricism. Both theories and the generative mechanisms they describe are real. Each generative mechanism has particular causal powers that are activated under specific conditions, and may produce quite different outcomes for different conditions (e.g., conditions can determine whether a system implementation project is a success or a failure). The theory of the mechanism, inferred via retroductive reasoning, postulates relationships, structures, and causal powers that are capable of producing a given effect, and when that effect will be triggered.

The third principle is that *critical realism distinguishes between an event (e.g., a decision or outcome) and our perception of that event.* For example, the rating of a seller on an online store is distinct from the rater's perception of the event, and also distinct from how anyone else sees or otherwise experiences the event. Data about that event gathered via a questionnaire is, of course, also distinct from the event itself. Our perceptions and the data that we access concerning an event are, in critical realism terminology, empirical traces. In the same way that events produced by a generative mechanism may differ from each other, empirical traces may differ from one occurrence of an event to another as may different traces of even the *same* event.

The fourth principle is that *critical realism uses an iterative and retroductive reasoning process as the basis for theory development*. Our understanding of the underlying cause of an event is always provisional because our reasoning process is fallible. Under the retroductive mode of reasoning, the key question posed when formulating a theory about the nature of the process that leads to a given of event is "what must the world be like for this phenomenon to be possible." As the discipline evolves, some theories will be corroborated, and others will be abandoned. Work by Hovorka (Hovorka et al. 2012; Hovorka et al. 2013) shows, in a striking visual format, the myriad tests of relationships between concepts in the IS field. Within the theory domains identified, a number of relationships show regular evidentiary support, while others have failed to achieve such support. Similarly, Hirschheim and Klein (2012) describe changes in how information systems development issues have been conceptualised and investigated throughout the history of the IS discipline, including the realisation that development issues can be more readily explained using a sociological perspective than is possible with the more technically-focused perspective used in the 1970s.

Principle 1	Critical realism allows a distinction between generative mechanisms and the actual events that they cause in particular circumstances.
Principle 2	Critical realism allows a distinction between a theory and the generative mechanisms (causal influences) that the theory describes.
Principle 3	Critical realism allows a distinction between the actual events and the empirical traces of these events which we can observe or experience.
Principle 4	Critical realism requires an iterative research process based on retroductive reasoning in which the central question is "what must the world be like for this phenomenon to be possible"

Table 1. Principles of critical realism for discussing research validity

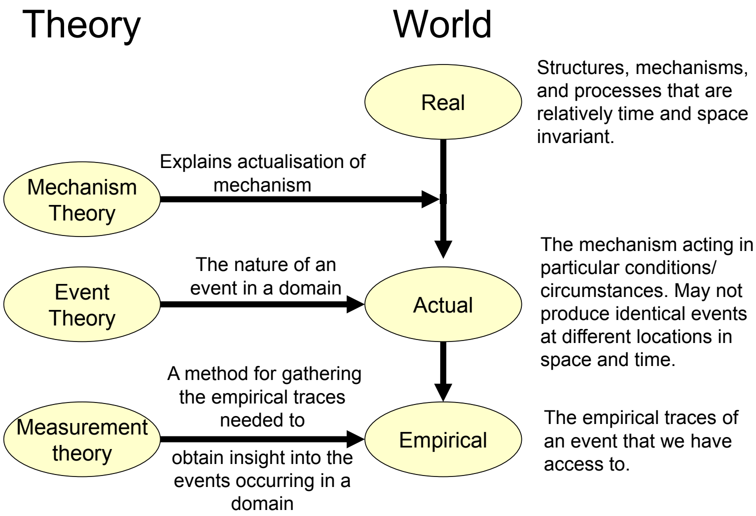

Figure 3. The relationship between theory and the three domains of the real

4.2 How these principles inform theory testing

These principles require that the role of theory testing is to assess the extent to which a theoretical account is consistent with what is known about a phenomenon observed in the IS problem domain, and indeed, the set of possible actual events in the domain of interest. However, because generative mechanisms, events, and empirical traces of those events are ontologically distinct, three types of theory (often conflated) must be distinguished when describing theory testing using a realist stance: a mechanism theory, an event theory, and a measurement theory. The relationship between each theory type and the operation of a mechanism in the world is illustrated in figure 3. Each is described in turn and explicated, first referring to *user acceptance of IT* concepts, and then to planetary motion (from the domain of cosmology) to show how the same theoretical structure is broadly applicable across both the natural and social sciences.

Mechanism theory

Mechanism theory proposes an explanation for how a mechanism is actualised as an event. For example, in the *user acceptance of IT* domain, researchers would like to explain the generative mechanism responsible for acceptance, and so explain all possible instantiations of the acceptance mechanism. The theory is derived, using a retroductive reasoning process, from a consideration of what the world must be like for the user acceptance to be possible. For example, each user acceptance theory is always subject to revision because the acceptance generative mechanism itself is unobservable and the theory itself is the product of fallible human reasoning, but the objective is always to explain a relationship between the real and actual layers. This can be stated more specifically as: mechanism theory is a testable description of how a given generative mechanism described can be triggered to cause the actual events that constitute the research problem domain to which the theory applies (See figures 2 and 3).

The approximate equivalent in cosmology to our user acceptance example is that the task of a theory of gravity is primarily to show how a certain conception of gravity explains the planetary orbits, not the position of the points of light we see in the sky (although it does that too). The empirical observations of the celestial positions of the planets is how we gain access to the actual orbits.

Event theory

An event theory describes the nature of events in a domain. This type of theory does not specify a specific measure or indicator, but instead characteristics of an event and how they give insight into the process that unfolds to cause the event. For example, a *user acceptance of IT* event theory would describe the characteristics of acceptance.

Like other types of theory, as the discipline evolves, some event theories will survive while others are abandoned. Marakas et al. (2007) describe this process in their account of how, as our understanding of underlying mechanisms has matured, our conceptualisation of the nature of the *computer self-efficacy* construct has changed along with our understanding of how it gives us access to IT usage phenomena. In cosmology, the equivalent development is the gradual refinement of theory describing the nature and properties of orbits, and how those properties reveal aspects of the operation of the underlying mechanism. Specifically, Tycho Brahe's empirical data on planetary trajectories in the night sky was the basis for Johannes Kepler's painstaking work in which he developed his three laws of planetary motion. Those laws describe the characteristics of orbits according to mathematical principles, and are a cornerstone of Newton's theory of gravity. In terms of our discussion, Newtonian gravity is a mechanism theory, while Kepler's work, which describes the characteristics of planetary motion, is the event theory.

Measurement theory

The relation between events and their empirical traces is the focus of the final type of theory, measurement theory. In the stratified ontology of critical realism, empirical traces are ontologically distinct from the events that produce them. This type of theory, therefore, proposes a method for gathering the empirical traces (data) needed to obtain insight into the events occurring in a domain. Unfortunately, for those wishing to conduct research, empirical traces are not neatly labelled and clearly bound to a single event. So a trace can have different meanings (or no meaning) depending on the context. For example, a standard measure of perceived usefulness "Using the application would increase my productivity" (based on Davis 1989, p. 340), potentially has no meaning outside of the largely utilitarian work context. A measurement theory should explain the relationship between context and measure, and the limits of that relationship, so that the meaning of a given trace in the particular context of the study is clear. Through the cumulative tradition of research into a topic, some measurement theories will be abandoned, and others will be enhanced. Marakas et al. (2007) describe a number of ways in which measures of computer self-efficacy have improved over time to describe the concept more effectively, but also show that the efficacy of some measures has degraded over time. The underlying concept exists independently of the conceptualisation, but the historical transformation has rendered some measures, or at least aspects of them, inadequate.

In the cosmology discipline example, initially the measurement theory was a set of geometric techniques for using earth-bound directional information on celestial bodies to reconstruct a three-dimensional orbit. However, new measurement techniques have been introduced recently, such as the use of Doppler redshift to assess the distance and velocity of distant celestial objects, and of lasers to track the lunar orbit. Visual recording has been supplanted by the new measurement techniques because it is conceptually and empirically inadequate. Underpinning each method is a theory of measurement that explains how we make use of specific empirical traces to provide information on the events of interest.

4.3 Implications of critical realism for validity

How theory is developed and tested, as implied by critical realism's stratified ontology, therefore differs in important ways from those considered valid under the empiricist view of the world. One important consequence of these differences is that, to some extent at least, concepts of validity in research are turned on their head compared to the empiricist's view.

Measurement validity

Measurement validity refers to how well a measure gives information about the thing it is designed to measure. For an empiricist, measurement validity refers to the statistical characteristics of set of observations about the construct (e.g., variance, skewness, kurtosis, and so on) and the connection between a measure and a theoretical concept. This form of validity is assessed using statistical methods such as the multitrait-multimethod matrix (Campbell and Fiske 1959), which calculates convergence and discriminance scores, and factor analysis (Spearman 1904), which uses the clustering of correlations of responses as a way to assess consistency with a theoretical concept.

For a critical realist, empirical qualities are only important insofar as they give information about the actual events occurring in the laboratory. In other words, validity is not primarily a property of the data, but a logical assessment of the relationship between data and event. Making a measurement validity claim involves assembling a chain of evidence about the quality of information the measurement gives about events. As a consequence, measurement validity is not just concerned with purely empirical issues, such as whether the data is biased, but also more conceptual issues, such as whether the measure is adequate to detect traces of the event.

Internal validity

Internal validity, for the empiricist, refers to whether observed changes can be uniquely attributed to the research intervention (other influences are accounted for primarily by exclusion from the setting, randomisation, or measurement) in which a number of treatments or stimuli are presented, and responses or results from each compared. Quoting Cook and Campbell's classic definition (1979, p. 37), it "refers to the approximate validity with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause." This form of validity is assessed using a *deductive* process on the basis of three criteria (Cook and Campbell 1979):

1. the presumed causal event precedes the effect recorded (temporal precedence),
2. the cause and the effect are statistically related (via covariation), and
3. there is no plausible alternative explanation for the conjunction of events.

For the critical realist, generative mechanisms are ontologically distinct from the patterns of events, so internal validity is concerned with more than events and their conjunctions. Internal validity, on this view, is a *retroductive analysis of* whether the study allows the researcher to identify the mode of operation of natural structures, mechanisms, or processes (that the researcher does *not* produce) using events and measurements (that the researcher *does* produce) (see Bhaskar 1998b, pp. 9-11). That is, if the objects of research activity are structures and generative mechanisms, how meticulously the researcher controls the events in the laboratory is a necessary but not sufficient basis for determining the internal validity of a study, because the research procedures and events produced alone do not make research activities intelligible. For the critical realist, therefore, internal validity is concerned with establishing a chain of evidence that the generative mechanism is a possible cause of the events occurring in the study or laboratory. When the experimental procedure has high internal validity, we are able to distinguish between events and the operation of the structures, mechanisms, or processes that produce the events, and which are the real objects of the experimental investigation.

Ecological validity

The term ecological validity, as used in the information systems discipline, refers to the representativeness of the design, or whether the setting of an experiment mimics the look and feel of a natural environment in the specific domain of practice being studied (see Bronfenbrenner 1977; Kelly et al. 2010; Vance et al. 2008). Used in this sense, ecological validity is really only relevant for investigations of environment-specific behaviours such as we find in a practice-oriented discipline like IS. For example, people do not, as a matter of course, draw UML diagrams of databases. Because that behaviour only happens in quite specific conditions, a study will be regarded has being ecologically valid to the extent that the closed system constructed to study how people construct UML diagrams resembles the sort of open system found in practice.

Critical realism, by making a higher-level ontological distinction between causal mechanisms and events, views this type of validity rather differently. Making this distinction allows us to recognise that executing a study in a naturalistic setting is not done to support an inference that a given constant conjunction of data exists in the natural setting modelled. Instead, when investigating the generative mechanisms that determine outcomes related to practice, the environment should be as naturalistic as possible to help ensure the created environment is not either activating a different generative mechanism to what we would expect to find in practice, or activating the right generative mechanism in an atypical way.

Ecological validity, when viewed through this lens, is the strength of evidence that the theorised mechanism for a phenomenon in the domain of practice was activated in the research environment in a way that is reflective of practice. Ecological validity is therefore related to internal validity, and can greatly enhance internal validity, although a study can show internal validity without demonstrating ecological validity.

External validity

Finally, although for the empiricist, external validity is a contentious concept, it is nevertheless generally conceptualised, following Cook and Campbell (1979, p.37), as "the approximate validity with which we can infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings, and times." In a sense, it is an assessment of how readily a presumed cause-effect relationship found can be extrapolated to the set of studies that we did *not* execute. When a study has no external validity, the effect found is specific to that study, making it impossible to generalize from the study's specific measures of particular observations made in particular circumstances to other measures and other situations.

External validity is commonly assessed in terms of statistical bias in sample descriptive variables. For example, Jiang et al. (2000) in a validation of the SERVQual instrument claim that their findings have acceptable external validity because there is no systematic bias in the size of organisations surveyed, the distribution of SERVQual gap responses is not biased, and an analysis of variance calculations found no systematic bias from any demographic variable, including age, gender, managerial position, and experience. From that analysis, the authors conclude that the SERVQual metric appears to capture user perceptions of service quality accurately and can therefore provide to managers deep insight into firm performance across multiple dimensions of service quality.

From a critical realist perspective, such a conceptualisation is bogus, because it conflates the empirical traces with the event, and the event with the mechanism. The critical realist view of external validity, therefore, is that it concerns the evidence that the generative mechanism that caused the actual events in the research setting also operates outside of the specific domain of the research. Evidence to support external validity claims includes being able to explain previously unexplained actual events beyond the research setting via the proposed generative mechanism.

Ecological validity is sometimes confused with external validity. However, in critical realist terms, external validity deals with the operation of the generative mechanism in contexts that were not studied, whereas ecological validity deals with the manner in which the generative mechanism relevant to a particular practice was activated.

The overall objective of assessing these four forms of validity, for the critical realist, is to establish a chain of evidence that the theorised generative mechanism can be said to cause the actual events in the problem domain and the empirical traces purportedly generated by those events. If we can establish all four forms of validity, we can be said to have tested theory, which is to show that we assessed the operation of the generative mechanism as described in our theory, and can therefore explain the operation of that mechanism in situations outside of the specific (laboratory) setting of the study. These issues can also be extended to research approaches that are not laboratory-based (such as a survey, field experiment, or positivist case study) by interpreting the term *laboratory* to mean the bounded and (semi-)controlled domain of empirical data collection. The conceptual differences between the empiricist and critical realist interpretations of validity issues is summarised in table 2.

	Empiricism	*Critical realism*
Measurement validity	Whether the statistical properties of the measure show that the measure is unbiased related to similar measures, and unrelated to measures that should be dissimilar.	Evidence that the measurement gives information about the events of interest.
Internal validity	Whether observed changes can be attributed to the research intervention rather than other possible influences.	Evidence that the particular proposed generative mechanism and only that mechanism has been triggered to cause the event/effect observed.
Ecological validity	The extent to which the environment constructed (the methods, materials and setting) corresponds to a natural setting outside of the laboratory where the effect being examined is believed to occur.	Evidence that the theorised mechanism for a phenomenon in the domain of practice was activated in the research environment in a way that is reflective of practice.
External validity	The approximate validity with which we can infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings, and times.	Evidence that the mechanism that causes the events in the research environment causes events outside of the specific domain of the research.

Table 2. Concepts of validity under empiricism and critical realism

It is particularly worth noting the new clarity that this formulation gives to external validity. On this view devising a research procedure to empirically test research hypotheses derived from theory is merely a stratagem for manifesting the purported generative mechanism in the controlled environments of the laboratory. But the task of showing that this generative mechanism also explains the actual events in the broader research domain (often incorrectly referred to as the *real world*) is no longer primarily concerned with the limited empirical traces of these actual events as they might manifest in that broader domain, but rather, about showing that the broad phenomena to be explained by the research are actually explained by the same mechanism that were invoked in a contrived and controlled way in the laboratory.

5 Exemplifying the validity issues

To exemplify our principles, assume that a researcher notices that whenever an experienced programmer is hired to work on a mobile device application for a fixed fee, the quality of work seems to vary according to whether the programmer is paid before or after delivery. The researcher then engages in a retroductive reasoning process, based on an understanding of practice, in which theory is developed to answer the question "what must the world be like for programing quality differences to manifest according to the payment terms of a contract?" Note that the meaning of quality here is based in the understanding and practice of the discipline, and the wider problem also necessarily refers to practice and indeed is part of the social process of knowledge development in the community of practitioners (Sayer 1992, pp. 14-15) Through this process, the researcher eventually theorises that the operation of an individual's internal and unobservable effort-reward system (the real mechanism) is affected by the method used to pay for the work.

To test this theory, the researcher designs a controlled experiment to compare outcomes from a group that is prepaid with one that is paid after completion. This design includes a measure of whether the trigger occurred (e.g., that the payment method is understood) and procedures to disambiguate the event from other parts of the environment. This experiment proceeds as follows:

1. We design a development task that requires use of the software, and two payment methods.
2. We take two groups of programmers with approximately equal ability.
3. We explain to each person the task requirements.
4. We ask each person to complete the set task and pay according to the rules for that group.
5. We observe development behaviour and collect measures of behavioural outcomes for each programmer in each group so that we can assess whether the event of interest occurred (differences in program quality), and how it occurred.

Note that, according to critical realism, observing which events occur and determining *why* is the main point of the exercise, *not* testing hypotheses *per se*. In addition, having *a priori* theory is not critical except that, in experimental research, one cannot normally ensure that the environment will trigger the event of interest without a strong theoretical base. Recording observations first and then finding they are explained by a theory is not as definitive as predicting something new and then observing it (hypothesis-testing), but quite common in the physical sciences (including geology, evolutionary biology, and cosmology). Regardless of the point at which theory is developed, validity issues must be addressed to make claims about the operation of the generative mechanism, particularly issues concerning measurement validity, internal validity, ecological validity, and external validity. These will now be discussed.

5.1 Measurement validity

In the terminology of critical realism, measurement validity is an assessment of whether the partial and perspectival empirical data collected under research conditions is an empirical trace of the actual events of research interest that operate below the empirical surface. Assessing measurement validity in this social research context requires evidence of a valid connection between what can be observed empirically (say the number of interactions required to perform a task) and what the information systems community agrees (consensually) to be the actual experience.

The key feature of this formulation of measurement validity is that it relates empirical traces to real events rather than to theoretical ideas. An important insight is that since the domain of actual events is underspecified by their empirical traces, measurement validity cannot be established simply by statistical analysis of the empirical traces themselves. Other procedures must be employed to demonstrate the connection of what is measured to what is occurring is the separate ontological domain of the actual, that is, to uncover the meaning of the data for actual events.

For the specific case we are examining here, we want at least one indicator of programming quality, potentially more if quality is a complex and multidimensional concept. Some indicators may be quantitative (e.g., number of interactions required to perform a task). For these measures it is reasonable that each quantitative measure be assessed using standard construct development practices that have emerged from the community of practice of researchers in our field (e.g., Lewis et al. 2005). Other indicators may be verbal or visual, collected by observing the types of activities performed during development and asking our surrogate employees about particular attitudes. These activities will create a line of evidence so that we have confidence that we are measuring programming quality.

5.2 Internal validity

Internal validity means establishing that the actual events that occur in the controlled domain are caused only by the generative mechanism that the theory proposes. The putative generative mechanism our hypothetical researcher proposes is that the effort-reward system is triggered by the method of payment on the fixed price contract (whether it is before work commences or after completion). The data gathered, particularly manipulation checks, should allow us to assess whether the laboratory events are consistent with the operation of this generative mechanism, at least in a well-defined statistical sense. Indeed, the patterns of data will have been predicted ahead of time and then triangulated by the measures of objective behaviour, attitudes (quantitative) and perceptions (verbal). We still, of course, need to eliminate explanations involving alternative generative mechanisms (e.g., prior experience, and so on) so that we can establish whether differences in outcomes are observed as a result of the operation of the effort-reward system.

Bear in mind, however, that in critical realism terms, the manipulation of the payment method is at the empirical level. This manipulation, therefore, is not the actual variable that is involved in the dynamics of the effort-reward system; it merely instantiates the initial conditions required for that generative mechanism to play out its dynamics. If the conditions are set in one way, these dynamics should lead to a lower mean quality of product produced. Further, the real entity that is the concern of the study, the effort-reward system, is not observable directly but real, although we can observe some aspects of specific instances of events, and it can perhaps be manipulated by changing the payment method. This is why part of the task of demonstrating internal validity must be to show that the manipulation controls the variable that triggers the mechanism.

5.3 Ecological validity

Ecological validity refers to the strength of evidence that the mechanism that causes the events in the research environment is the same as the theorised mechanism for the phenomenon as it occurs in the natural open systems within the practical problem domain. For the programming domain problem, providing a naturalistic work area, requirements documentation, and work contract may make the environment appear natural to the experimenter, but it is evidence from the participants that ultimately matters.

5.4 External validity

External validity means that the generative mechanism that explains events within the constructed environment of the study also explains phenomena outside of the study, particularly in the broader domain of practice about which research questions are usually formulated. In the context of our case, external validity means assembling evidence that the putative effort-reward system proposed is found in a natural environment and can explain other outcomes. For example, we could look at other environments where an effort-reward system governs behaviour, and the how our effort-reward mechanism can explain outcomes in those other environments.

This last point is crucial and is a point that seems to be missed often by researchers in our discipline, resulting in ultimately irrelevant studies (sometimes even published in top journals) that are *internally* valid but not *externally* valid. It is also the motivation behind some extended discourses on this problem including (Lee and Baskerville 2003; Seddon and Scheepers (2006). Because our research is human-centered research, we must ensure that our design allows us to extrapolate beyond the study to humans in the general domain of practice, otherwise the effect of the generative mechanism may not be the same. More importantly the mechanism may not be the same. For example, if the group recruited has quite different values to the typical experienced professional programmer, the effort-reward system may operate quite differently. That is, if the mechanism of interest is activated, the outcomes are unique to that group (for example, the post-paid group might produce programs that are only activated once payment is received). Therefore, recruitment (and screening) procedures are vital, as are more subtle issues, such as ensuring that each participant has the required skills.

6 Conclusion

In this paper our aim was to show that new clarity can be brought to the formulation of the traditional validity issues of theory testing research in IS by explaining the fundamental principles of each issue in the language of critical realism. The separation of the vague and undifferentiated notion of theory into the three specific types of theory we have described clarifies the research process by showing how each is necessary to model the traces, events, structures, and causal mechanisms present in the different layers of the ontology. Our reformulation of the four types of validity described provides clarity regarding meaning and assessment of the traditional validity issues relevant to theory testing in practice-oriented IS research.

Applying critical realism to each validity issue has provided important new insights. Measurement validity concerns whether the partial and perspectival empirical data collected under research conditions are generated by, and describe in an adequate form (for the purposes of the research), the actual events of interest that operate below the empirical surface. Consequently the researcher must be concerned with establishing that the data collected represent the underlying actual event variables both structurally and meaningfully. Since the category of the actual is distinct and larger that the empirical this cannot be established entirely by analysis of empirical data, but is now seen as a theory-laden process (note that this gives a theoretical justification to construct validation procedures already described in the IS literature, such as Lewis et al. (2005)).

Internal validity means establishing that the actual events that occur in the controlled domain are caused by the generative mechanism that the theory proposes. This requires the researcher to compare the events which follow from the purported generative with those revealed by the empirical data as in the hypothetico-deductive method, but also to ensure alternative possible generative mechanisms for the events are eliminated.

Ecological validity is the strength of evidence that the theorised mechanism for a phenomenon in the domain of practice was activated in the research environment in a way that is reflective of practice, for example, by setting the study in an environment that mimics the natural setting of the proposed generative mechanism. This requires the researcher to construct both the setting and experimental controls so that the generative mechanism of interest is allowed to operate while excluding (or nullifying) the effect of other mechanisms that could affect results.

External validity involves establishing that the generative mechanism that explains events within the study also causes phenomena outside of that setting, particularly in the broader domain of practice about which research questions are usually formulated. This means that the research manipulations that occur in the controlled research situation should invoke the same generative mechanism that produces the phenomena we wish to explain.

Our systematic exploration of the consequences of critical realism for validity provides several novel contributions to theory over those derived from empiricism, which tends to inform formal discussion of these issues. External validity is essentially undefined under hard empiricism since the latter does not support a concept of an external world of practice to which the concept refers. This has led IS researchers to supplement empiricism in *ad hoc* ways with realist notions, with confusing results. The notion from critical realism of generative mechanisms, and the new interpretation of theory testing, brings new clarity to this issue and validity in general. The notion that the generative mechanisms that are invoked in contrived research situations must equate to those that operate under ambient circumstance is particularly novel and important. It provides a new razor by which research can be judged as relevant and important to practice.

References

Archer, M.S., (2013). *Social origins of educational systems*, Routledge, London.

Bealer, G., (1992). The Incoherence of Empiricism. *Proceedings of the Aristotelian Society*, (66): 99-138.

Bhaskar, R., (1978). *A Realist Theory of Science*, (2nd ed.), Alma Book Company, York.

Bhaskar, R., (1993). *Dialectic: The Pulse of Freedom*, Verso, London.

Bhaskar, R., (1998a). *Critical Realism*, Routledge, New York.

Bhaskar, R., (1998b). *The possibility of naturalism: A philosophical critique of the contemporary human sciences*, (3rd ed.), Routledge, London.

Boudreau, M.C., Gefen, D., and Straub, D.W., (2001). Validation in Information Systems research: A state-of-the-art assessment. *MIS Quarterly*, (25:1): 1-16.

Bronfenbrenner, U., (1977). Toward an experimental ecology of human development. *American psychologist*, (32:7): 513-531.

Bunge, M., (2006). *Chasing Reality: Strife Over Realism*, University of Toronto Press, Toronto.

Campbell, D.T., and Fiske, D.W., (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, (56:2): 81-105.

Carlsson, S.A., (2003). Critical Realism: A way forward in IS research. in: *The 11th European Conference on Information Systems (ECIS'2003),* C. Ciborra, R. Mercurio, E.D. Marco, M. Martinez and A. Carignani (eds.), Naples, Italy. 1-12.

Carlsson, S.A., Henningsson, S., Hratstinski, S., and Keller, C., (2010). Socio-Technical IS Design Science Research: Developing Design Theory for IS. *Information Systems and e-Business Management*, (9:1): 109-131.

Collier, A., (2011). The social ontology of critical realism. In: *Sociological Realism,* A. Maccarini, E. Morandi and R. Prandini (eds.), Routledge, London, pp. 3-20.

Cook, T.D., and Campbell, D.T., (1979). *Quasi-experimentation: Design & Analysis Issues for Field Settings*, Houghton Mifflin Company, New York.

Davis, F.D., (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technologies. *MIS Quarterly*, (13:3): 319–340.

Dobson, P.J., (2002). Critical Realism and Information Systems research: why bother with philosophy. *Information Research*, (7:2): Paper 124.

Fogelin, R.J., (1993). Hume's scepticism. In: *The Cambridge Companion to Hume,* D.F. Norton (ed.), Cambridge University Press, Cambridge, pp. 64-89.

Greenwood, J.D., (1990). Two dogmas of Neo-Empiricism: The "theory-informity" of observation and the Quine-Duhem thesis. *Philosophy of Science*, (57:4): 553-574.

Hirschheim, R., and Klein, H.K., (2012). A glorious and not-so-short history of the Information Systems field. *Journal of the Association for Information Systems*, (13:4): 188-235.

Hooker, C.A., (1975). Philosophy and meta-philosophy of science: Empiricism, Popperianism and Realism. *Synthese*, (32:1): 177-231.

Hovorka, D.S., Birt, J., Larsen, K.R., and Finnie, G., (2012). Visualizing the core-periphery distinction in theory domains. In: Proceedings of the 23rd Australasian Conference on Information Systems (ACIS 2012), ACIS, pp. 1-10.

Hovorka, D.S., Larsen, K.R., Birt, J., and Finnie, G., (2013). A meta-theoretic approach to theory integration in Information Systems. In: Proceedings of The 46th Hawaii International Conference on System Sciences (HICSS 2013), IEEE, pp. 4656-4665.

Jiang, J.J., Klein, G., and Crampton, S.M., (2000). A note on SERVQUAL reliability and validity in information system service quality measurement. *Decision Sciences*, (31:3): 725-744.

Kelly, D., Fu, X., and Shah, C., (2010). Effects of position and number of relevant documents retrieved on users' evaluations of system performance. *ACM Transactions on Information Systems (TOIS)*, (28:2): Article 9.

King, W.R., and He, J., (2005). External validity in IS survey research. *Communications of the Association for Information Systems*, (16): Article 45.

Klein, H.K., and Lyytinen, K., (1985). The poverty of Scientism in Information Systems. In: *Research methods in Information Systems,* E. Mumford, R. Hirschheim, G. Fitzgerald and T. Wood-Harper (eds.), Elsevier Science, Amsterdam, pp. 131-161.

Klein, H.K., and Myers, M.D., (1999). A set of principles for conducting and evaluating interpretive field studies in Information Systems. *MIS Quarterly*, (23:1): 67-94.

Kolokowski, L., (1972). *Positivist Philosophy: from Hume to the Vienna Circle*, Penguin Books, Harmondsworth.

Lee, A.S., and Baskerville, R.L., (2003). Generalizing Generalizability in Information Systems Research. *Information Systems Research*, (14:3): 221-243.

Lewis, B.R., Templeton, G.F., and Byrd, T.A., (2005). A methodology for construct development in MIS research. *European Journal of Information Systems*, (14:4): 388-400.

Lucas, J.W., (2003). Theory-Testing, Generalization, and the Problem of External Validity. *Sociological Theory*, (21:3): 236-253.

Maddy, P., (1992). Indispensability and Practice. *Journal of Philosophy*, (89:6): 275-289.

Madill, A., Jordan, A., and Shirley, C., (2000). Objectivity and reliability in qualitative analysis: Realist, contextualist and radical constructionist epistemologies. *British Journal of Psychology*, (91:1): 1-20.

Marakas, G.M., Johnson, R.D., and Clay, P.F., (2007). The Evolving Nature of the Computer Self-Efficacy Construct: An Empirical Investigation of Measurement Construction, Validity, Reliability and Stability Over Time. *Journal of the Association for Information Systems*, (8:1): Article 2.

Mingers, J., (2000). The contribution of Critical Realism as an underpinning philosophy for OR/MS and systems. *Journal of the Operational Research Society*, (51:11): 1256-1270.

Mingers, J., (2003). A critique of statistical modelling from a Critical Realist perspective. in: *The 11th European Conference on Information Systems (ECIS'2003),* C. Ciborra, R. Mercurio, E.D. Marco, M. Martinez and A. Carignani (eds.), Naples, Italy. Paper 81.

Mingers, J., (2004a). Critical Realism and multimethodology. In: *Critical Realist Applications in Organisation and Management Studies,* S. Fleetwood and S. Ackroyd (eds.), Routledge, London.

Mingers, J., (2004b). Real-izing Information Systems: Critical Realism as an underpinning philosophy for Information Systems. *Information and Organization*, (14:2): 87-103.

Mingers, J., Mutch, A., and Willcocks, L., (2013). Critical Realism in Information Systems research. *MIS Quarterly*, (37:3): 795-802.

Neuman, W.L., (2005). *Social Research Methods: Qualitative and Quantitative Approaches*, (6th ed.), Allyn and Bacon, Milwaukee.

O'Leary-Kelly, S.W., and Vokurka, R.J., (1998). The empirical assessment of construct validity. *Journal of Operations Management*, (16:4): 387-405.

Popper, K.R., (1959). *The Logic of Scientific Discovery*, Hutchinson, London.

Rosenberg, A., (1993). Hume and the philosophy of science. In: *The Cambridge Companion to Hume,* D.F. Norton (ed.), Cambridge University Press, Cambridge, pp. 64-89.

Sayer, R.A., (1992). *Method in Social Science: A Realist Approach*, Routledge, London.

Smith, M.L., (2006). Overcoming theory-practice inconsistencies: Critical Realism and information systems research. *Information and Organization*, (16:3): 191-211.

Spearman, C., (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, (15:2): 201-292.

Teddon, P.B., and Scheepers, R., (2006). Other-settings generalization in IS research. in: *The 27th International Conference on Information Systems (ICIS 2006),* , Milwaukee, WI. 1141-1158.

Tsang, E.W.K., (2013). Case studies and generalization in Information Systems research: A Critical Realist perspective. *The Journal of Strategic Information Systems*: in press.

Vance, A., Elie-Dit-Cosaque, C., and Straub, D.W., (2008). Examining trust in information technology artifacts: the effects of system quality and culture. *Journal of Management Information Systems*, (24:4): 73-100.

Volkoff, O., and Strong, D.M., (2013). Critical Realism and affordances: Theorizing IT-associated organizational change processes. *MIS Quarterly*, (37:3): 819-834.

Williams, K., (1975). Facing reality—A critique of Karl Popper's empiricism. *Economy and Society*, (4:3): 309-358.

Wynn, D., Jr., and Williams, C.K., (2012). Principles for Conducting Critical Realist Case Study Research in Information Systems. *MIS Quarterly*, (36:3): 787-810.