2. How critical realism clarifies validity issues in theory-testing research: analysis and case

ROBERT B. JOHNSTON
UNIVERSITY COLLEGE DUBLIN

STEPHEN P. SMITH
MONASH UNIVERSITY AND UNIVERSITY OF MELBOURNE

Abstract

Much discussion of research validity in information systems (IS) draws on the empiricist tradition, which presents an impoverished account of reality, making it difficult to formulate issues of validity that are pertinent to a practitioner-oriented discipline. This is particularly true for external validity, which refers to the applicability, or relevance, of research to phenomena in the broader environment of practices, but also for construct and internal validity, which seek to establish rigour in human-centred practical research. Discussions of validity are often supplemented with implicitly realist notions that are not compatible with the espoused empiricism. We argue that critical realism presents a number of insights that profoundly clarify a discussion of validity in practitioner-oriented theory-testing research. We reformulate the notions of theory testing and research validity from the critical realist perspective and illustrate them using a description of a controlled experiment. Important new insights are gained into establishing each kind of validity.
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 and the amount of column space in more conventional theory-testing articles that purports to demonstrate that the work is valid in some way. In research that comes from the ‘positivist’ tradition in IS—such as laboratory experiments, surveys and theory-testing case studies—discussions of research validity typically draw either implicitly or explicitly on the empiricist position of quantitative social science for their justification. Because empiricism formulates the process of research only in terms of the domain of theory and the domain of empirical data and barely admits the existence of an independent ontological domain beyond the experience of the senses that we might identify with the ‘real world’, discussions of validity drawing on this tradition have difficulty formulating research validity issues in a way that would be appropriate for IS, which is principally a practice-based research domain. 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 practices, but it is also true of construct and internal validity that seek to establish the rigour of human-centred practical research in which many important constructs are socially constructed or experiential.

As a consequence, IS researchers tend to supplement the apparatus of empiricism with realist notions that are in fact incompatible with the ontology of hard empiricism (Mingers 2004b; Smith 2006). Discussions of research design issues for the IS field are therefore often quite opaque because the theoretical bases for them are neither entirely explicit nor completely consistent philosophically. We diagnose the root cause of this difficulty in discussing validity issues to the poverty of empiricism as a framework for these discussions and the uncontrolled way in which writers on research methodology try to compensate for this poverty with naive realist notions—or else with notions from ontologically incompatible constructivist or interpretivist traditions.

In this chapter, we want 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 (1975, 1993, 1998), Niiniluoto (1999), Sayer (1992), and others. Critical realism has an ontology that admits a number of useful categories with which to formulate validity issues, and additionally has an epistemology that recognises the socially constructed nature of both the socio-technical practices that are the focus of IS research and the IS practice itself. With critical realism, it is possible to distinguish between the theory
to be tested and the generative mechanisms to which the theory refers as causes of the events that are observed in both the contrived and the bounded circumstances of the research study (the ‘laboratory’), as well as in the broader ambient circumstance to which research questions are supposed to apply (the so-called ‘real world’ of practice). These new ontological categories and distinctions profoundly simplify and clarify a discussion of construct validity, internal validity and, particularly, external validity in theory-testing research. We will show this by a detailed exposition of validity issues from a critical realist position and exemplify the analysis by applying it to a real laboratory experiment conducted by the authors.

We understand that nowadays most IS researchers would reject the stronger forms of empiricism described below. Actual research practice and pedagogy provide recipes for ensuring validity, which move beyond that framework and include many of the insights that we find follow from a critical realist perspective (Mingers 2000; Smith 2006). Our point is that these extensions lack an adequate philosophical foundation as long as empiricism continues to be the framework used for their exposition. Consequently, precise analysis and justification for validity procedures cannot be given. This is especially true for external validity where empiricism provides little guidance. Our argument 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 those philosophies. 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 systematised in the following section. Building on these principles, we describe an actual 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 arguments and their broader implications.

The standard empiricist view of theory testing

Inquiry 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 (Hamlyn 1967; Hooker 1975). Hume’s scepticism—to the effect that no knowledge can be established conclusively by reason alone (and that even belief in an external world and/or the existence of the self are not justifiable)—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).

Positivism is an extreme extension of empiricism that rejects all claims of knowledge except those derived empirically. In 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 maximise 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), to a positivist, a statement that one event causes another is equivalent to saying that one event is always followed by another. As a result, the (highly influential) positivist view of scientific inquiry is that the proper role of theory is for predicting patterns based on objective observations (because causation cannot be proved).

Theory testing, from an empiricist standpoint, finds its ultimate expression in Popper's (1959) falsificationism, which holds that because we can never prove a theory definitively, a scientific hypothesis must be evaluated in terms of whether it is in fact false. Theory testing, in 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 2.1, which shows the relationship between the empirical and theoretical domains under that paradigm. Note that this figure shows 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 2004b; Smith 2006; Williams 1975), so are not the focus of our discussion. For the purposes of this discussion, however, two points should be noted regarding Figure 2.1. First, on a strict reading of empiricism, only empirical and theory categories enter the empiricism paradigm. Limiting the objects of inquiry 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).
2. How critical realism clarifies validity issues in theory-testing research

Second, limiting the research focus to a comparison of experience with theory means that construct validity, internal validity and external validity must all be concerned primarily with empirical issues. The relationship between the data and the laboratory for each of these forms of validity is illustrated in Figure 2.1 and explained next.

Construct validity, for the empiricist, is concerned with the relationship between data and concepts described in theory. In that view, construct validity is established by demonstrating that parts of the measure are consistently related but distinct from other measures. Cronbach and Meehl (1955:283) (who created the term) define a construct as a ‘postulated attribute of people assumed to be reflected in test performance’. To validate a construct, ‘a nomological net surrounding the construct must exist’ (Cronbach and Meehl 1955:291) so that the researcher can assess the strength of evidence that the label attached to a measure is an accurate description of the construct actually measured (for example, examine the correlation between the data collected and one or more variables that are believed to be related to the construct of interest). In other words, construct validity is the amount of correspondence between an unobservable concept and measurable indicators of that concept’s properties.
Because the emphasis of validation is on the strength of the evidence within a nomological network, assessment of validity, in research practice, tends to stress the statistical structure of the construct (especially coherence and differentiation) while neglecting issues to do with meaning.

*Internal validity* is concerned with showing that there is a ‘constant conjunction’ between the empirical events observed as patterns. For an experiment, this means demonstrating that the constructed environment forms a closed system by virtue of the experimental controls. To the extent that closure can be demonstrated, the observed event sequences can be attributed to the research intervention because alternative explanations will be untenable.

*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 1979:37). 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. Constructing any argument about the validity of assumptions outside the specific data set examined (or other known data sets) ultimately requires, however, 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 centred conception of validity are reflected in the obscurity of discussion of this issue in positivist-oriented IS literature. For example, Lee and Baskerville’s (2003) lengthy discussion about non-statistical forms of generalisability essentially proposes pragmatic work-arounds to some of the more 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 above, it is clear that construct validity is problematic in IS research because many constructs we would like to measure have socially constructed or experiential aspects, and there is no way to assess the correspondence between a measure and a behaviour or other observable phenomenon. Moreover, in circumstances in which it is possible to observe an outcome (for example, technology use), because of the close relationship between data, measures and constructs, empiricists tend to conflate a trait with the test for that trait, resulting in theoretical properties of a construct being defined (at least partially) in terms of data values (Bunge 2006).
Internal validity is relatively straightforward to achieve in experimental studies. In a practice-oriented research field, however, it is highly problematic for many research methods, including theory-driven surveys, case studies and action research. In non-manipulative research, the only evidence of constant conjunction is the rather weak similarity between observed and predicted association, which provides little reason to believe that one variable is causally related to the other. Because the environment is not a closed system, demonstrating internal validity, in these cases, really requires additional evidence that the observation is not a chance association—for example, via time series data.

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 generalisations 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 mathematics, IS and other disciplines indicate widespread departures from strict empiricist requirements, creating an implicit tension between ‘ideal’ and actual practice (see Bealer 1992; Maddy 1992; Smith 2006). In IS research in particular, these differences indicate that IS researchers already view aspects of the research process in transcendental realist terms (Mingers 2004b; Smith 2006), so critical realism could be a more accurate formulation of what researchers really believe and practice. We argue that conceptualising validity only in terms of the relationship between theory and observation has led to a great deal of confusion about validity issues in IS research—particularly external validity—because IS researchers tacitly deviate from the espoused empiricist tradition in realist directions (Smith 2006), so are unable to explain some research practices in strict empiricist terms. The more sophisticated realist position we expound in the next section illuminates this issue and provides a set of conceptual tools that helps 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.
Brief outline of critical realism

Bhaskar’s (1993, 1998) philosophy of critical realism—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 (reality) and that only some aspects of this world are objectively knowable via our senses. Our senses are not always completely reliable, of course—for example, we can be fooled by illusions and we can misinterpret sense data. Nevertheless, because reality is independent of our senses, 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 operates even if we are not aware of its operation.

Bhaskar (1975) distinguishes between transitive and intransitive objects of knowledge in the world. Intransitive objects are the ‘real things and structures, mechanisms and processes, events and possibilities of the world; and for the most part they are quite independent of us’ (Bhaskar 1975:22). That is, the existence of an intransitive object does not depend on our knowledge or perception of it. Transitive objects, on the other hand, include theories, paradigms, models and methods. These objects are subjective and their existence is dependent on human activity (if people suddenly ceased to exist, transitive objects would cease to exist).

These distinctions 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 views on 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—an anthropocentric concept—is what people can experience. Things in the actual world, such as events, cannot be perceived, but events leave empirical traces, and we can perceive those traces. That is, the sensing of an event is not the same as the event (which can occur regardless of whether we can sense it). In addition, limitations on our senses mean that we might not perceive all traces (that is, of the things in the empirical domain that we can experience, we will experience only a subset), and the subjective and perspectival nature of perception means that experiences will vary from one person to another and from one setting to another.

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) refers to generative mechanisms as alethic truths—the underlying processes that
give rise to both actual and empirical events and the phenomena that scientists seek to identify. Explaining this concept, Groff (2000:411) writes, 'To be the alethic truth (y) of x is to be the causal structure, or generative mechanism, that gives rise to x'. 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 non-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 instantiations of the generative mechanisms). Finally, experiences are obtained via empirical traces of actual things, so 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 2.2, derived from Mingers (2004b:94, Figure 1).

![Figure 2.2 The three domains of the real (stratified ontology of critical realism)](image)

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, 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, in 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 researcher could manipulate the environment surrounding a particular class of decision process to answer the questions why someone would 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.

Having covered the essential characteristics of generative mechanisms (and other real entities), it should be noted at this point that empiricists hold that causality is a relation between two events, but explain why particular events are related though dispositions held by objects (a trait exhibited only when certain conditions that obtain or realise that property occur). A standard example is that a wine glass is fragile, but this property is not realised ordinarily, only when an enabling circumstance occurs, such as when the glass is dropped (Singer 1954). So the empiricist’s idea of a disposition is a bit like a generative mechanism. Because all relationships between events are disposition based, there is, however, no conceptual terminology available to an empiricist to distinguish between a test for an event and properties of that event, leading to an almost inevitable conflation of a trait with the test for that trait—for example, saying that a property of an acid is that it turns blue litmus paper red (Bunge 2006:244).

With this background, we can now see how this philosophical orientation applies to objects of knowledge in the natural sciences, in which an event and empirical traces of that event can be explained and predicted by a generative mechanism. The proper job of the natural scientist, according to Bhaskar, is to attempt systematically to identify the entities responsible for an event and to describe the generative mechanism. He argues (Bhaskar 1975:195–6) that these principles also apply to the social sciences, in which the objects of inquiry are socially constructed entities, such as social structures (Durkheim’s major work *Suicide* [1897] is cited as an exemplar for the social sciences), but also that the transitive nature of the structures examined in the social sciences (compared with 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 localised in time and space and hold only for particular contexts—and epistemologically, social science is self-referential in the sense that, like the phenomena studied, it is itself a social practice (Mingers 2004a). Bhaskar argues, however, that these structures nevertheless exist independently of the researcher and the research activity and it is in that sense that the real, actual and empirical domains can still be examined. Furthermore, the fact that the structures are dependent on humans does not make them unreal or any less worthy of study. For example, languages are clearly transitive, but we also know that they are separate from the people who use them. Their structures, changes in them and even their very existence
are types of events that must have been produced by generative mechanisms. The job of the social scientist is also therefore to explain the structures, objects and generative mechanisms.

### Implications of critical realism for theory testing in IS

The threefold stratified ‘naturalist’ ontology just presented differs substantially from positivism and interpretivism—the dominant espoused research paradigms in the IS 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 principles of critical realism in the context of IS research, and with reference to Figure 2.3, and show how these principles assist our understanding of validity issues in research. A description of how these principles were applied in a particular research project follows and is summarised in Table 2.1.

First, critical realism distinguishes between theory and the generative mechanism that the theory describes. Both theories and the generative mechanisms that they describe are real. Theories are transitive and are located in the social practice of science. Generative mechanisms are intransitive and located in the domain to which the theory applies, which in turn can be partly natural and partly social. The distinction between a theory and the generative mechanism to which it refers is, however, the key to understanding how the implications of critical realism for theory testing and validity differ from those of empiricism.

Strictly speaking, IS has few, if any, theories that can be said to truly describe a generative mechanism (at least, there is nothing comparable with widely applicable theories in the natural sciences, such as Einstein’s theory of relativity). Most theories, instead, describe regularities between research entities located within the *actual* domain. For example, many pre-existing theories have been used to explain the phenomenon of technology adoption (including diffusion theory [Rogers 1995], the theory of planned behaviour [Ajzen 1991], and so on), while others have been developed specifically for that purpose (including the Technology Acceptance Model [Davis 1989] and its variants). These theories have particular built-in assumptions, apply from particular perspectives (for example, end-user, developer, manager) or apply under particular circumstances (for example, voluntary versus mandatory use), so are intended to explain *particular* instances of the phenomenon, rather than all possible occurrences.
This leads to the second principle: that critical realism distinguishes the actual consequences of the generative mechanism in specific circumstances from the generative mechanism itself (because, as shown in Figure 2.2, the actual is an instantiation of the mechanism in the real). So an actual consequence—say, user acceptance in a particular situation (employment or personal use)—is different from the generative mechanism underlying that outcome.

The third principle is that critical realism distinguishes between an actual instance (what happened) and empirical traces of the actual (our perception of what happened). For example, voluntary acceptance by a user of a personal organiser is distinct from the empirical traces via which we can obtain knowledge of that process (measures of ease of use, usefulness, intention, and so on).

Theory testing, in this view (and principle four), is concerned with showing that the generative mechanism explains phenomena observed in the problem domain (which is broader than explaining the empirical). In critical realism terms, this can be stated more specifically: theory testing means showing that the generative mechanism that the theory describes produces the actual events that constitute the research domain to which the theory applies (see Figures 2.2 and 2.3). For example, in the user acceptance of IT domain, we as researchers would like to explain the generative mechanism responsible for phenomena in the user acceptance domain, so explain all possible instantiations of that mechanism,
which can be experienced via empirical traces. In other words, theory testing is concerned with systematically connecting the posited generative mechanism to the set of possible actual events in the domain of interest. The approximate equivalent in physics is that the task of a theory of relativity is primarily to explain that 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. In contrast with the empiricist position, however, critical realism asserts that this access to the actual is not entirely via the senses but also via theory (which is a significant point in the later section when we discuss construct validity). This focus on explaining the actual (the relationship between a generative mechanism and an actual instantiation) is radically distinct from hard empiricism in which it is merely the empirical that needs to be explained, and it is largely for this reason that critical realism views empiricism as an impoverished research philosophy.

Table 2.1 Principles of critical realism for discussing research validity

| Principle 1 | Critical realism allows a distinction between a theory and the generative mechanisms (causal influences) that the theory describes |
| Principle 2 | Critical realism allows a distinction between generative mechanisms and the particular events that they cause in particular circumstances |
| Principle 3 | Critical realism allows a distinction between the actual events we would like to explain and the empirical traces of these events that we can observe |
| Principle 4 | Under critical realism, theory testing is showing that the generative mechanism that the theory describes produces the actual events that constitute the research domain to which the theory applies |

Implications of critical realism for validity

The research approaches implied by the threefold stratified ontology therefore differ in important ways from those considered valid under the empiricist view of the world. Moreover, because the focus of the research becomes the generative mechanism underlying events (rather than the observations of the events), to some extent at least, concepts of validity in research are turned on their head compared with the empiricist’s view. For an empiricist, construct validity refers to the correspondence between observations and the theoretical construct, whereas for a critical realist, construct validity is concerned with whether empirical traces give information about the actual events occurring in the laboratory that are purportedly caused by the generative mechanism (that is, whether the observations are a manifestation of the actual phenomenon of interest). Internal validity for the empiricist refers to whether observed changes can be attributed to the research intervention rather than other possible influences (primarily by excluding or measuring the effect of other factors that
might produce the effect examined). For the critical realist, on the other hand, internal validity is concerned with establishing that the generative mechanism is the cause of the actual events observed in the study or laboratory. Finally, although for the empiricist external validity is a contentious concept, it is nevertheless generally conceptualised as the extent to which the relationships found in the data can be generalised to other types of people, settings and times (that is, generalised from specific observations to events in other situations). 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 represents the likelihood that the generative mechanism that caused the actual events in the laboratory also causes the events that occur more widely in the problem domain (within the broad research boundary). These distinctions are summarised in Table 2.2.

**Table 2.2 Concepts of validity in empiricism and critical realism**

<table>
<thead>
<tr>
<th>Construct validity</th>
<th>Empiricism</th>
<th>Critical realism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Whether the measure is consistent with the theoretical description of the construct.</td>
<td>Whether data that are empirically available give valid knowledge about the actual manifestation of the purported generative mechanism in the laboratory.</td>
</tr>
<tr>
<td>Internal validity</td>
<td>Whether observed changes can be attributed to the research intervention rather than other possible influences.</td>
<td>Whether actual events are manifestations of the particular generative mechanism in the laboratory circumstance.</td>
</tr>
<tr>
<td>External validity</td>
<td>Whether the researcher can infer that the presumed causal relationship generalises to and across alternative measures of the cause and effect and across different types of people, settings and times.</td>
<td>The likelihood that similar or related events that occur (or might occur) in other settings are caused by the generative mechanism that caused the actual events in the laboratory.</td>
</tr>
</tbody>
</table>

The overall objective of assessing these three forms of validity for the critical realist is to establish the extent to which the generative mechanism can be said to cause the actual events in the problem domain, with each form of validity operating in a different space within the three domains of the real. Construct validity establishes that the data (which are all that we have direct sensory access to) that are empirically available give valid knowledge about the actual manifestation of the purported real generative mechanism in the laboratory. Internal validity establishes that these actual events are a manifestation of the particular generative mechanism under laboratory circumstances. Finally, external validity establishes that actual events in the broader research boundary can be accurately predicted from the proposed generative mechanism. Note that 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.

It is also worth noting the new clarity that this formulation gives to external validity. In this view, forming hypotheses to be empirically tested 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 concerned primarily with the limited empirical traces of these actual events as they might manifest in that broader domain, but rather with 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.

**Case illustration of application of validity principles**

To exemplify how the application of critical realism principles can strengthen IS research, we refer to a study carried out by the authors (Smith et al. 2008) on consumer responses to product evaluation support technologies. This discussion of the research adopts a critical realist perspective. Using that study as a reference point, we illustrate the methods that might in fact be required to establish validity under this new view.

**Overview of study**

The specific problem analysed in the study was the impact on the product evaluation phase of a purchase transaction of a class of consumer-oriented software tools, called Virtual Model technology, with which a consumer could build a virtual self and then change the appearance of that self in a virtual dressing room (for example, change their hairstyle or try on virtual shirts, jeans, sunglasses and other wearable items). Consumer responses to different implementations of this concept were compared with responses to the more traditional catalogue-style interface to answer the specific question of whether customers who used a virtual model interface while evaluating felt more informed and whether feeling informed about a product led to a more positive attitude towards the online store.

In posing that question, we asked implicitly whether providing increasingly sophisticated technology was the best way to deliver effective customer support online, arguing that the technological determinism assumed by that development approach was untenable. Based on established theories from
the consumer decision-making literature, we predicted that how a consumer perceived the product being evaluated (the object of the task) would cause that person to actively seek different information. In other words, we predicted that the outcome of the evaluation task would be determined by the fit between the type of information provided and the type of information sought by the consumer, rather than the fit between the technology and the task (as task-based theories commonly cited in the IS literature would suggest, including cognitive fit [Vessey and Galletta 1991] and the Task-Technology Fit model [Goodhue and Thompson 1995]).

The investigation of these ideas was operationalised via a controlled experiment in which participants assessed products online using either a basic catalogue or a description supplemented with a virtual model display. Unrelated systems at two live e-stores were used in the experiment and responses were collected from evaluations of four separate products. Verbal responses throughout each evaluation were recorded, providing rich accounts of attitudes and thoughts, and each participant also completed a survey to provide quantitative data about the experience.

Critical realism perspective of research

Using this brief description of the research method and aims, we are now able to illustrate the application of our argument to this research situation and to discuss how validity as conceived from the critical realist conception can be achieved through research design in practice. The explanation in the paragraphs following mirrors the content of Figure 2.4.

In the centre bottom of Figure 2.4, we show two entities: 1) the informational affordance of the web store (its potential to provide information of a certain kind); and 2) the information-seeking attitude of the user (which, we argue, determines the information they will find useful). Because, from a critical realist perspective, information is real (Mingers 2004b), the informational affordance of the web store is an intransitive real thing. It is, however, a potential that is actualised only in a certain circumstance—namely, when it is encountered by a user seeking that type of information. On the other hand, the information-seeking attitude of the user is entirely subjective, so would be classified as a transitive entity. Despite this, it would be viewed as real by critical realism because it has the potential to cause empirically observable effects (clicking the ‘order’ button, for example). The information affordance of the web site and the user attitude are not, however, actual in themselves because they do not actually produce a specific effect until certain circumstances obtain—namely, that the end-user with a particular attitude to the product encounters the web-store display that provides information relevant to that attitude. Consequently,
these two entities are part of the generative mechanism that produces the actual effects that we predict will be empirically observed in the laboratory and occur more widely in the broader research domain (online apparel shopping).

Figure 2.4 Research constructs and validity for the research example in Smith et al. (2008)

At least, this is what the theory that we developed and that we are testing proposes, and that theory itself is an actual instantiation of more general theories in the domain of consumer behaviour research. This theory domain is real because it has observable effects, such as the production of our previous research papers. It is a real domain of human practice and thus transitive. Critical realism, however, distinguishes between our actual theory and the generative mechanism to which it refers, the actual effects of that generative mechanism in particular circumstances (the user’s experiences) and empirical traces of these events observed in the controlled situation (the questionnaire responses and taped oral responses of the experimental subjects).

According to our argument, the objective of testing our theory is to collect evidence that the generative mechanism to which the theory refers explains the events that occur in the broader domain to which our theory applies—namely, online apparel shopping. The research stratagem that our experiment employs is to make predictions about the nature of the empirical traces that should be observed of events that occur in rather controlled circumstances to carefully selected potential customers under careful manipulation of the web sites they encounter. We make these predictions by a process of logical deduction.
about how our purported generative mechanism will actualise itself under the conditions in the laboratory (hypothesis generation) and how we would expect the empirical traces of these events to be observed (operationalisation). Finally, we seek to design the experimental situation so that the conclusions we draw from comparing the empirical traces produced in the experiment with those predicted from our theory validly allow us to have confidence that the mechanism that produced this agreement in the controlled situation also explains phenomena that regularly occur in the broader domain of the research, and thus to answer the research question ‘Under what circumstances, and why, would providing Virtual-Model technology to consumers lead them to use an e-store?’

**Exemplifying the validity issues**

A critical realist perspective of our research question is that a specific type of event—a customer having an experience with a product information display in the field—arises from the putative mechanism described in our theory. Observing this event and explaining the underlying causal mechanism is the point of the research. To do so, however, we need to ensure that requisite conditions for triggering that event are present and also that we can disambiguate the trigger and event from other parts of the environment. To evaluate this problem, therefore, we perform an experiment that (in idealised form) proceeds as follows:

- we take two information resources that provide different types of information
- we take two groups of users who each seek one of the two types of information
- we show them each type of information
- we ask each person to use that information to evaluate a product as though considering buying that item
- we collect (self-reported) answers to predefined questions from each person that are designed to reveal whether the event of interest occurred.

It is important to note that testing hypotheses is not the main point of the exercise, but rather to observe what events occur and determine why. In addition, having a priori theory is also not critical except that, in experimental research, it is hard to ensure that the environment will trigger the event of interest or to observe the event and triggering conditions without a strong theoretical base. For example, we could record observations about an event and later find it is predicted by a theory (or even develop the theory). That approach is not as definitive as predicting something new and then observing it (hypothesis testing), but is quite common in the physical sciences (including geology, evolutionary biology and cosmology). Regardless of the point at which
theory is developed, being able to make any claims about the operation of the generative mechanism requires addressing validity issues, particularly ones concerning construct validity, internal validity and external validity. These will now be discussed.

Construct validity

In the terminology of critical realism, construct validity is an assessment of whether the partial and perspectival empirical data collected under research conditions are empirical traces of the actual events of research interest that operate below the empirical surface. Both the empirical data and the actual events are considered real. In contrast, under strict empiricism, construct validity describes the connection between the empirical data collected (the measure) and the theoretical conception of the construct. When the actual events occur in the natural world (for instance, inventory is moved from one place to another), relating empirical traces to actual events is a rather straightforward problem of physical measurement. Many of the events we wish to observe in IS research are, however, either social constructions of the relevant communities of practitioners (such as agile software development practices [Meso and Jain 2006]) or experiences of individual people (such as ‘feeling informed’, in our case). In both these cases, critical realism still holds that the events are real but that now they are in the transitive domain of the social. In the first case, we need to establish that what we measure corresponds with actual occurrences within the community of practice that produces and uses the social construction, and in the second case that what we measure gives an indication of a certain mental state within a particular person (subject), which we consensually agree to be a certain experience. Consequently, establishing construct validity in these cases will require establishing a valid connection between what can be observed empirically (say, an answer to an interview question) and what some community agrees (consensually) to be the actual event, occurrence or experience. The community might be practitioners, consumers, the general public or researchers, depending on the nature of the event being measured.

The key difference in this formulation of construct validity from that under empiricism is that here empirical traces are related to real events (perhaps socially constructed or experiential events) rather than to theoretical ideas. An important insight is that since the domain of actual events is under-specified by their empirical traces, construct 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 in 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 a measure of how informed someone feels about an item after examining relevant product information, but we also want to ensure that we gather data that help us to assess the meaning of that measure—did we measure the right thing? We performed standard construct validation procedures on all experiential measures used in the survey by asking surrogate consumers to group rate and group propose questions in relation to the experience the question purported to measure. That created a line of evidence that when a certain empirical trace (a certain answer to a question) was observed, the underlying event had been experienced by the subject. We also asked the subject to talk aloud about what was experienced when a given product representation was viewed. This provided a second and independent empirical trace (speech, language) of the actual experience of the subject—albeit one with its own validity issues that were of a largely hermeneutic nature and that were addressed with standard content analysis techniques.

Internal validity

Internal validity means establishing that the actual events that occur in the controlled domain are caused by the generative mechanism that the theory proposes. That is, internal validity is established through: 1) explanation of the mechanism; 2) confirmation that the mechanism has operated as described; and 3) elimination of alternative explanations (for example, experiment procedures and measures that point to other possible causal mechanisms). In other words, testing is explanation. Testing can involve prediction, but not necessarily—particularly if the objective is historical analysis. It is worth noting that under this revised conceptualisation of internal validity, theory testing can employ a strategy similar to the hypothetico-deductive method, but the chain of logic does not have to be from theory to data to be valid; rather, provided the generative mechanism can be explained, assessment might in fact be validly driven by data. It is common in formal research practice to seek explanations from data in this way even when controlled experiments are performed, but because of the influence of empiricism in IS, research is usually written up as though hypotheses were derived from theory before observations. This is because empiricism requires prediction to precede observation to eliminate the possibility that the regularities observed in data might be spurious. In critical realism, identification of a consistent generative mechanism can provide an alternative guarantee.

In terms of our case, a defensible association between the event observed and the causal mechanism must be given to demonstrate internal validity. The generative mechanism we proposed was the fit between the information provided by a given style of product representation and the type of information the subject was seeking given their attitude to the product. The events we
observed (as indicated by both the patterns of answering the survey questions and the verbal description of the subjects’ own experiences) were shown to be consistent with the operation of this generative mechanism at least in a well-defined statistical sense. Indeed, the patterns of survey answers were predicted ahead of time and triangulated by the self-reporting data. We still, of course, need to eliminate explanations involving alternative generative mechanisms (for example, controlling treatment order, eliminating brand-recognition bias, recruiting consumers who have not used the systems previously, and so on). Because the aim is to explain the actual, if, however, we observe the result that we predicted, we also need to establish whether it occurred as a result of the mechanism described.

Bear in mind, however, that in critical realism terms, the manipulations are at the empirical level. Manipulations, therefore, are not the actual variables that are involved in the dynamics of the generative mechanism; they merely instantiate the initial conditions required for the generative mechanism to play out its dynamics. If the conditions are set in one way, these dynamics should lead to the consumer feeling informed. Set another way, they will lead to a different feeling (perhaps indifference). Further, the real entity that is the concern of the study—*information affordance*—is not directly observable through the real, but can be manipulated by changing the web interface viewed. This is why part of the task of demonstrating internal validity must be to show that the manipulation controls the variable: in our case, that to present the virtual model is, in fact, to make available a certain kind of informational affordance.

**External validity**

External validity means establishing that the generative mechanism that explains events in the ‘laboratory’ is also the generative mechanism that causes the actual phenomena of the broader domain of practice about which research questions are usually formulated. This means the researcher must establish that the event assessed (experimental treatment) invokes the same generative mechanism that operates in the ambient situation.

To achieve this goal, both the treatments and people (and other parts of the laboratory environment) should be good surrogates for the real-world phenomenon that we are claiming to investigate. This is the key to external validity—a point that seems to be missed often by researchers in our discipline, which accounts for many of the ultimately irrelevant studies that have been published that are *internally* valid but not *externally* valid and which is the motivation behind some extended discourses on this problem, including Lee and Baskerville (2003) and Seddon and Scheepers (2006).
In the context of our case, external validity means creating a controlled situation that invokes the same generative mechanism that is found in the real world. For example, given the choice between developing technologies ourselves and using live web sites, external validity considerations suggest building treatments around live sites (if we developed our own, we would not be sure that responses were governed by the same mechanism). Other considerations include whether the products examined are representative of the products that are available online (again, the use of live stores addresses this issue), whether the laboratory task itself is representative and whether the people are representative of the general population of Internet shoppers who could be subject to the generative mechanism of interest. This last point is crucial. Because our research is human centred, we must ensure that the humans involved are close surrogates for the users of the technology under study, otherwise the effect of the generative mechanism might not be the same. Therefore, recruitment (and screening) procedures are vital, as are more subtle population-related issues, such as ensuring that each participant evaluates only products that are personally relevant (for example, people who do not wear prescription eyewear should not evaluate eyeglass frames, only people who actually wear jeans should evaluate jeans, and so on) and the motivation of the participants during the study (many experiments involving decision making have no real consequences for the decisions taken, which is very different from the real world).

The point of reference in these considerations is that, conceptually, the environment that we create for the purposes of the experiment is a controlled form of the world, in terms not just of populations, but of the causal mechanisms that are at work at any given time. We therefore must ensure correspondence between what is done in the laboratory world with what is done outside it, in terms of both the population involved and the tasks performed. By ignoring underlying causal mechanisms, empiricism tends to direct attention to the former at the expense of the latter.

**Final discussion**

Overall, therefore, if we can establish these three types of validity, we can be said to have tested theory, which is to show that we assessed the operation of the generative mechanism described in our theory and can therefore explain the operation of that mechanism in situations outside the specific (laboratory) setting of the study. Theory testing, in this view, differs markedly from requirements informed by empiricism. In that view, we would need to show that: 1) the treatment design operationalises the theoretical construct—the information affordance of the resource; 2) the information-seeking attitude of the user varies independently of the treatment design (there is no demand effect); and 3) the measure operationalises the feeling informed construct. We would then need to
show that our design had excluded potential confounds (internal validity) and the presence of a statistically significant association between the manipulations and the outcomes (in the manner predicted). Ironically, the actual event of being informed (the focus of the study) does not enter into this prescription. Moreover, it cannot enter into our considerations because it is not observable and is thus not an empirical entity. But, more importantly, because empiricism allows theory to be formulated in terms of expected values of the concepts, the actual event is redundant at the theory level.

The difference in the critical realism requirements—described under the various forms of validity—and empiricist requirements is significant. First, critical realism sees construct validity as a relation between the dependent variable operationalisation and the actual event that is supposed to occur, not between data and theory. Second, internal validity is concerned with whether the generative mechanism is responsible for (actually causes) the actual event, which is indirectly detected from the empirical traces, not with the logic of theory deduction (although this is important). Third, external validity refers to whether the events in the experiment are caused by the same mechanism that causes the phenomenon of interest in the field and thus whether the experiment provides evidence that the putative generative mechanism causes the effects in the field, which is the aim of theory testing. The empirical level is really of relevance only in the laboratory where a test of predictions is the issue. Consequently, inferential statistics are *not* the major concern when assessing external validity.

**Conclusion**

In this chapter, we have reformulated the notions of theory testing and research validity from the perspective of critical realism. Compared with an impoverished empiricist account of validity for practitioner research, critical realism brings important new ontological categories and distinctions to the problem. These include the notion of generative mechanisms distinct from theories, the distinction between generative mechanisms and their actual effects in particular circumstances and the distinction between these actual effects and the empirical traces that we observe.

Applying critical realism to each key validity issue has provided important new insights. Construct 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 variables both structurally and meaningfully. Since the category of the actual is distinct and larger than the empirical, this cannot be established entirely by analysis of empirical data, but is now seen as a theory-laden process.

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 that follow from the purported generative mechanism 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.

External validity involves establishing that the generative mechanism that explains events in the ‘laboratory’ is also the generative mechanism that causes the actual phenomena of 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. Furthermore, since many variables that constitute these events are socially constructed or human centred, subjects must respond to the generative mechanisms in ways similar to practitioners.

Our systematic exploration of the consequences of critical realism for validity provides several novel contributions to theory compared with those derived from empiricism alone. 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 bring new clarity to this issue and validity in general. The notion that the generative mechanisms that are invoked in contrived research situations must equate with 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.

In addition, critical realism underlines the limited extent to which research validity can be established by statistical manipulation of data alone. Empirical data give limited information about only the actual events that occur in the contrived or bounded research situation. Accessing the domains of actual events and real generative mechanisms is an essentially theory-bound process. Similarly, although comparing empirical data collected with predictions is a valid internal validity stratagem, external validity is not at heart an issue of statistical generalisation from those data.
The significance of this work for research practice lies in the way that important principles and practical procedures for conducting valid theory-testing research can be systematically derived and justified from a well-developed realist philosophical framework that embraces the complexities of socio-technical practice-based research.

References


2. How critical realism clarifies validity issues in theory-testing research


