A Scientific Basis for Rigor and Relevance in Information Systems Research

Allen S. Lee

School of Business, Virginia Commonwealth University

Richmond, Virginia 23284-4000, USA

allenslee@alum.mit.edu

+1 804.827.0366

Geoffrey S. Hubona

Robinson College of Business, Georgia State University

Atlanta, Georgia 30301

+1 404.413.7364

25 April 2008

This paper is being distributed to participants in the Friday Research Workshop hosted by the Information and Decision Sciences Department and the MISRC at the University of Minnesota on 2 May 2008. Please do not make copies of this paper or distribute it in any form without the authors' permission.

A Scientific Basis for Rigor and Relevance in Information-Systems Research

Acknowledgments

Sergey Samoilenko, in his role as research assistant in 2005, motivated the first author to articulate his long-held, tacit ideas about the logic of empirical inquiry in science. He began "beta testing" these ideas by presenting them in research seminars in numerous venues: the Information Technology Postgraduate Education Program for IT-Professionals of Western Finland and Oulu Provinces, the action-research pre-conference workshop at AMCIS in Omaha, the University of Colorado, Simon Fraser University, Pepperdine University, Indiana University, the University of Central Florida, the University of North Carolina at Greensboro, the College of William & Mary, Virginia Commonwealth University, Queen's University of Belfast, the London School of Economics, James Madison University, the *MIS Quarterly* New Authors Workshop at ICIS in Milwaukee and in Montreal, Washington State University, Virginia Polytechnic Institute and State University, and the University of Arizona. Comments from Jeffrey V. Nickerson of the Stevens Institute of Technology and Damien Fennell of the London School of Economics further contributed to the development of the arguments in this essay.

A Scientific Basis for Rigor and Relevance in Information-Systems Research

Abstract

Qualitative research is just as able as quantitative research to follow certain fundamental principles of logic in general and scientific reasoning in particular. Two such principles are the logic of *modus ponens* and the logic of *modus tollens*. In this essay, we frame different research approaches – positivist research, interpretive research, action research, and design research – in the forms of modus ponens and modus tollens. Three issues emerge from this framing and call into question how research is now conducted in the discipline of information systems. They are the issue of a common scientific basis, the issue of the fallacy of affirming the consequent, and the issue of summative validity. Both rigor and relevance in information-systems research may be better achieved by attending to the three issues.

A Scientific Basis for Rigor and Relevance in Information-Systems Research

For many years, the academic discipline of information systems has considered qualitative research to be no less valid than quantitative research (Markus 1997). The road to acceptance traveled by qualitative research offers a lesson worth revisiting. The lesson is that qualitative research is just as able as quantitative research to follow certain fundamental principles of logic in general and scientific reasoning in particular. In this essay, "formal logic" or simply "logic" will refer to the ways by which a researcher relates abstract symbols and propositions to one another, and "science," which incorporates logic, will additionally refer to the ways by which a researcher relates the symbols and propositions to empirical referents.[1]

As a proponent of qualitative research, this essay's first author has explained how qualitative research is able to follow these principles, where he has offered explanations in the form of narrative (Lee 1989; Lee 1991; Lee 1999; Lee and Baskerville 2003). Still, because he originally derived his ideas about scientific reasoning from some fundamentals of formal logic, he has always believed that he could more effectively convey his ideas if he were to express them in the notation of formal logic. He has finally done this (see Appendices A and B). This essay will examine the consequences that these ideas, expressed in formal notation, have for empirical inquiry in information-systems research. As is often the case when framing ideas in mathematics or other formal notation, the notation has had the effect of sharpening the ideas and leading to new, unanticipated issues. In fact, the ramifications for quantitative and positivist

---

[1] Propositions and symbols, as used in formal logic, need not have any empirical content or referents. For example, algebra can be viewed as involving the logic of how to relate mathematical propositions and symbols to one another, where applications of them to the "real world" need not be made. Non-Euclidean geometries provide an example of propositions and symbols that are not even meant to have empirical referents in the first place. Logic itself is not empirical. In contrast, propositions and symbols, as used in science, need to have empirical referents in the "real world." In scientific research that is positivist, the task of relating a theory's propositions and symbols to the real world is typically pursued through the rules of experimental and quasi-experimental design, whereas formal logic has no need for these rules.

research[2] may even be more significant that the ramifications for qualitative and interpretive research.[3]

We are raising three methodological issues in this research essay: the issue of a common scientific basis, the issue of the fallacy of affirming the consequent, and the issue of summative validity. The issues can be controversial and call into question how information-systems

---

[2] In the philosophy of science and the social sciences, the term "positivism" has numerous meanings, not all of which are consistent with one another. For example, positivism, as introduced by August Comte, is different from the logical positivism embraced by the Vienna Circle. And though sometimes identified as part of the positivist Vienna Circle, Karl Popper regarded himself as antipositivist. Furthermore, within the philosophy of science, positivism can be considered discredited; Schön quotes Bernstein (1976, p. 207): "There is not a single major thesis advanced by either nineteenth century Positivists or the Vienna Circle that has not been devastatingly criticized when measured by the Positivists' own standards for philosophical argument. The original formulations of the analytic-synthetic dichotomy and the verifiability criterion of meaning have been abandoned. It has been effectively shown that the Positivists' understanding of the natural sciences and the formal disciplines is grossly oversimplified. Whatever one's final judgment about the current disputes in the post-empiricist philosophy and history of science … there is rational agreement about the inadequacy of the original Positivist understanding of science, knowledge and meaning." Two helpful monographs on positivism in philosophy are provided by Bernstein (1976) and Kolakowski (1968).

Positivism in social science is influenced by positivism in philosophy, but also distinct from it, especially considering that the philosophy of science would consider any inquiry relying on the discredited verifiability criterion of meaning to be incorrect or even infeasible in the first place. As used in this essay, "positivism" refers to a genre of social-science research that, among other things, regards the natural sciences as the model for the social sciences to live up to. In subscribing to the "natural science model," positivism puts forth elements often associated with the natural sciences – e.g., independent and dependent variables, mathematical propositions, quantitative data, inferential statistics, and experimental controls – which it requires the social sciences to incorporate if they are to become as scientific as the natural sciences. This description of the term "positivism" is consistent with its usage among information-systems researchers including Orlkowski and Baroudi (1991), Ngwenyama and Lee (1997), Trauth and Jessup (2000), Dubé and Paré (2003), and Weber (2004).

[3] "Interpretive research," as used in this essay, is associated with ethnography, hermeneutics, and some forms of case research. It acknowledges that the understanding held by the human subjects in the researcher's field of study (i.e., the "subjective understanding") is part and parcel of the overall subject matter that the researcher is studying, and therefore requires observation and data collection no less than any other part of the objectively existing subject matter being studied. In interpreting what the local setting and its context mean from the "natives' point of view," the researcher develops an "interpretive understanding" of the subjective understanding. Fundamentals of interpretive research are provided by Lee (1991), Orlikowski & Baroudi (1991), and Walsham (1995).

research is now practiced. To support our argument, we will use a framework that we build from some elementary aspects of formal logic. We call it the MPMT framework, where MPMT refers to a specific way of using *modus ponens* and *modus tollens*.

The main contribution we intend to make in this essay is to show how the MPMT framework provides a scientific basis for the rigor of research, where the bulk of our examination focuses on rigor in positivist research and interpretive research. A corollary to this examination will be that the MPMT framework can just as well provide a scientific basis for the rigor of research which focuses on relevance, such as action research and design research.[4]

The MPMT Framework

Modus ponens and modus tollens are well established and accepted forms of syllogistic reasoning. The deductive logic of the syllogism is a fundamental of formal logic and is covered in textbooks used in introductory philosophy and logic courses. As such, modus ponens and modus tollens provide a sound framework with which to identify the ramifications that formal logic can have for the reasoning used in scientific research.

Syllogistic reasoning is perhaps best known in the following form: “All humans are mortal,” “Socrates is a human,” and “therefore, Socrates is mortal,” where the statements are instances of the syllogism’s major premise, minor premise, and conclusion, respectively. As straightforward as this reasoning may seem, the straightforwardness is deceptive. A general familiarity with syllogistic reasoning does not substitute for a detailed understanding of the major premise, minor premise, and conclusion as they are used in the empirical testing of a scientific theory.

Theories can take diverse forms (Gregor, 2006). Not all theories (including positivist theories) do, can, should, or must fit the MPMT framework, including the framework’s mathematical notation (introduced below). However, in the process of building a theory, a

---

[4] We use the term “design research” as a synonym for “design science.”

researcher can make the choice of shaping it to fit the MPMT framework, whereupon the researcher could use the framework as a basis on which to establish the theory's rigor. In this essay, we will provide examples of how this is accomplished.

Because positivism is familiar to a majority of information-systems researchers, it is useful as a starting point for illustrating the logic of the MPMT framework. The following illustration will focus on positivist behavioral research in information systems that involves statistical inference and multivariate analysis. Published positivist behavioral research in information systems often provides a boxes-and-arrows or ellipses-and-arrows diagram as a depiction of a theory's variables and the relationships among them. This visual depiction, in turn, corresponds to a mathematical depiction in which there is an equation for each dependent variable, the general form of which is, "$Y_i=f(X_{i1}, X_{i2,} \ldots ,X_{in})$ for $i=1,2,\ldots,m$," where n is the number of independent variables and m is the number of equations or dependent variables.

As helpful as mathematical notation can be, however, it must be noted that a theory is necessarily more than just any mathematical representation of it. Any such mathematical representation exists in the context of other elements of the theory, such as verbally (not mathematically) expressed assumptions, conditions, and definitions, as well as the empirical referents that the theory purports to be about. Without these other elements, the mathematical symbols and equations – what Kuhn (1962, 1977) calls "symbolic generalizations" – would literally be meaningless. Kuhn goes on to explain that there are also social and cognitive aspects of a scientific community which enable its members to have a shared understanding of the standard symbolic generalizations used in their community, but which also hinder or block the achievement of this understanding by members of a different scientific community. For the authors of this essay, a purpose of a mathematical representation of a theory is not to re-present every element of the theory, but to operationalize the theory into a form that would allow it to be empirically tested with the tools of statistical inference.

Not all positivist theories need be expressible in the form of ellipses-and-arrows diagrams and operationalized in the form of mathematical equations, but those positivist theories which can be expressed and operationalized in this fashion include many, if not most, of the behavioral theories that undergo statistical analysis in information-systems research. For example, these diagrams appear in studies by Davis, Bagozzi, and Warshaw (1989), Ang and Straub (1998), and Zhu and Kraemer (2005). Table 1 provides the corresponding mathematical equations (Kuhn's "symbolic generalizations") for the ellipses-and-arrows diagrams in the three studies. The general form "$Y_i=f(X_{i1}, X_{i2}, \ldots, X_{in})$ for i=1,2,…,m" covers linear, nonlinear (including polynomial), and interactive (moderating) relationships between variables and can be expressed in matrix notation (Mendenhall and Sincich, 2003, pp. 721-753). Representing a theory's variables and relationships in the form of equations has the benefit of allowing the use of powerful, multivariate statistical estimation procedures, such as multiple regression and structural equation modeling.

Table 2 describes an application of syllogistic logic in science. The sense in which Table 2 uses the term "prediction" requires explanation, as does another term in Table 2, "initial conditions." In the Notes for Table 2, we specify the technical meanings of these terms and explain the meanings of the symbols Y, X_i, x_i, $ß_i$, b_i, x_{new}, y_{new}, and $y_{predicted}$, which we use in the following illustrations.[5]

In Table 2, a theory whose variables and relationships are mathematically operationalized as "$Y = f(X_1, X_2, \ldots, X_n)$" (which, as a simplified form of the general case "$Y_i=f(X_{i1}, X_{i2}, \ldots, X_{in})$ for i=1, 2, … , m," is sufficient for purposes of illustration) is applied to the initial conditions ("$X_1=x_1$, $X_2= x_2$, $X_3= x_3, \ldots, X_n= x_n$"), which are the data or facts describing a particular experiment, organization, laboratory, population or other setting either before, or without, what

---

[5] In accordance with some statistical conventions, we use an upper-case letter to designate a random variable and a lower-case letter to designate an actual numerical value taken by the random value. The exceptions are *p* and *q*, which we use as statement variables in modus ponens and modus tollens.

positivist research describes as the application of the "experimental treatment," the "statistical treatment," or simply, the "treatment." The resulting conclusion is the prediction that the dependent variable, with or after the treatment's being administered, will take the numerical value "$y_{\text{predicted}}=f(x_1,x_2,x_3,\ldots,x_n)$." In positivist research involving theories that can be mathematically operationalized as "$Y = f(X_1, X_2, \ldots, X_n)$," this instance of the syllogism is a building block that supplies the major premise of both modus ponens and modus tollens.

Modus ponens is the form of the syllogism in which the major premise takes the form, "if *p* is true, then *q* is true"; the minor premise, "*p* is true"; and the conclusion, "therefore *q* is true" (the first column in Figure 1). An example of modus ponens is:

major premise:
If "all humans are mortal" is true, then "Socrates is mortal" is true.[6]

minor premise:
"All humans are mortal" is true.

conclusion:
Therefore, "Socrates is mortal" is true.

Modus tollens (the second column in Figure 1) takes a different form: the major premise is the same as in modus ponens, but the minor premise "*q* is not true" and the conclusion "therefore *p* is not true" are different. An example of modus tollens is:

major premise:
If "all humans are mortal" is true, then "Socrates is mortal" is true.

minor premise:
"Socrates is mortal" is not true.

conclusion:
Therefore, "all humans are mortal" is not true.

---

[6] The major premise here is the conclusion in the following conditional proof:

major premise: "All humans are mortal."
minor premise: "Socrates is a human."
conclusion: "Socrates is mortal."
"Therefore, if 'all humans are mortal' is true, then 'Socrates is mortal' is true."

When the statement "if *p* is true, then *q* is true" is used in the empirical inquiry of science instead of just logic in general,[7] *p* is a general or universal proposition denoting the theory and *q* is a singular proposition denoting facts or data that the researcher expects, if the theory is true, to observe in the particular experiment, organization, laboratory, population or other setting where the theory is being tested.[8] Tables 3 and 4 illustrate this in the case of positivist behavioral information-systems research involving theories that can be mathematically operationalized; here, *p* stands for the theory mathematically operationalized as "$Y = f(X_{1,} X_{2,} \dots , X_n)$" and *q* stands for the prediction "$y_{\text{predicted}}=f(x_1,x_2,x_3,\dots,x_n)$" which the theory makes in the particular setting where the theory is being tested. Later in this essay, we will illustrate *p* and *q* with examples from other forms of empirical inquiry: interpretive research, design research, and action research.

The incorporation of modus tollens as a feature of empirical reasoning (in contrast to logic in general) appears in the thinking of scholars as diverse as Argyris and Schön, Peirce, Copi, and Popper. In qualitative organizational studies, modus tollens can be found in Argyris and Schön's (1974) concepts of "testability" and "disconfirmability," which refer to whether or

---

[7] See footnote 1 for an explanation of the distinction between logic in general and the empirical inquiry of science.

[8] We can restate this with some formal-logic terminology. *p* and *q* in the major premise of modus ponens and modus tollens ("if *p* is true, then *q* is true") are bound by certain restrictions when used in the empirical inquiry of science: *p* is a "statement variable" that contains "individual variables" but no "individual constants"; *q* is the result of replacing the individual variables in *p* with individual constants; and individual constants correspond to specific empirical referents. In formal logic, individual variables can refer to, but are not limited to, random variables as used in statistics or to mathematical variables in general. An individual variable can refer to a set of things; "end user" is an individual variable and an end user who is survey respondent #1 would be not only an individual constant (the value taken by the variable), but also an empirical referent instantiating the individual variable "end user." Outside the empirical inquiry of science, these restrictions on *p* and *q* in modus ponens and modus tollens need not apply.

not a person's theory of action (*p*) is able to be corrected through the empirical testing of its consequences or predictions (*q*) in this person's everyday experience.[9] Argyris (1999) also uses "testability" and "disconfirmability" in his formulation of "action science." In the philosophy of pragmatism, whose founders are often considered to be William James, Charles S. Pierce, and John Dewey, modus tollens can be found in pragmatism's emphasis on the consequences (*q*) that follow from beliefs (*p*) about truth, rightness, or value – which cover beliefs not only in science, but in all areas of life.[10] In his textbook on logic, Copi (1986) explicitly uses modus tollens in his explanation of scientific testing, which involves the proposition which is tested (*p*) and also one or more other propositions (*q*), deduced from the proposition to be tested, that are capable of

---

[9] Argyris and Schön state (p. 25): "Theories of action are testable if one can specify the situation, the desired result, and the action through which the result is to be achieved. Testing consists of evaluating whether the action yields its predicted results. If it does, the theory has been confirmed; if it does not, it has been disconfirmed." Equivalently, the major premise "if *p* is true, then *q* is true" is "if 'the given theory of action' is true, then the prediction that 'the desired result occurs' is true." In the case of modus tollens, the minor premise "not *q*" is " 'the 'desired result' does not occur" and the conclusion "therefore not *p*" is "therefore 'the given theory of action' is not true." The conclusion would then motivate correction of and improvement in the given theory of action. Note that, for Argyris and Schön, the word "confirmed" means that the theory or action is consistent with observation, not that the theory is true.

[10] According to Rescher (2005, p. 747), if a person's belief (*p*) about the world has the quality of "truth in the case of statements, rightness in the case of actions, and value in the case of appraisals," then an action (*q*) based on it will "work out" (i.e., the major premise "if *p* is true, then *q* is true"). Then, suppose an action based on the belief does not "work out" (i.e., the minor premise, "not *q*"). Therefore the person's belief about the world would lack the quality of "truth in the case of statements, rightness in the case of actions, and value in the case of appraisals" (i.e., the conclusion, "not *p*"). This would then motivate a correction of and improvement in the person's belief, becoming p', which in turn would be empirically tested through its consequences, q'.

being tested directly.[11] Finally, Popper inserts modus tollens into his notion of “falsifiability” (1968a, pp. 54-56), where he puts forth “falsifiability as a criterion of demarcation” (p. 20) with which “to distinguish between the empirical sciences on the one hand, and mathematics and logic as well as ‘metaphysical’ systems on the other” (p. 11). Popper emphasized this by saying that falsifiability addresses the problem, “ ‘*When should a theory be ranked as scientific?*’ or ‘*Is there a criterion for the scientific character or status of a theory?*’,” where his intention was “*to distinguish between science and pseudo-science*” (1968b, p. 33, emphasis in the original). Whereas Popper’s overall philosophy has generated controversy, the validity of modus tollens has never been called into question and, furthermore, the recognition given to modus tollens in the empirical inquiry of science predates Popper’s use of it to demarcate science from non-science.[12]

We will use the MPMT framework to identify and examine three issues in information-systems research, where this framework appears not only in positivist research, just discussed, but also in interpretive research, where modus ponens and modus tollens take the forms detailed in Tables 5 and 6. These tables mention hermeneutic interpretation, which originated as the

---

[11] Under the heading of “Explanations: Scientific and Unscientific,” Copi states: “The pattern of indirect testing or indirect verification consists of two parts. First, one deduces from the proposition to be tested,” where this proposition is the theory *p*, “one or more other propositions capable of being tested directly,” which we recognize as *q*. Deducing *q* from *p* pertains to the major premise of modus tollens (if *p* is true, then *q* is true). “Then, these conclusions,” which are predictions, “are tested and are found to be either true or false. If the conclusions are false, any proposition that implies them must be false also. On the other hand, if the conclusions are true, that provides evidence for the truth of the proposition being tested, which is thus confirmed indirectly” (1986, p. 486).

[12] “…Popper advanced an idea that, though *not novel in the history of modern science,* was now formulated in all its generality with great clarity. This idea was that the empirical character of statements (and hence of their meaningfulness) should be their ‘defeasibility,’ that is, the possibility of disproving them” (Kolakowski, 1968, p. 185, emphasis added).

study of the interpretation of text, but extends to interpretivism in general where “text” is replaced by “text analogue” (see the Notes for Tables 5 and 6). The MPMT framework may readily incorporate the hermeneutic circle – which Klein and Myers (1999) describe as “fundamental” and “foundational” in interpretive field studies – as an instance of the logic of modus tollens (note f in the Notes for Tables 5 and 6).

The First Issue: a Common Scientific Basis

The first of three issues to emerge from the MPMT framework is that positivist research and interpretive research can build on a common scientific basis – namely, the application of modus tollens for the purpose of empirically testing a theory. Consider that the first column of Table 4 (for positivist research) and the first column of Table 6 (for interpretive research) *share the same formal-logic argument* – namely, modus tollens. This shows that the logic of modus tollens is blind to whether the propositions applying it happen to appear in positivist or interpretive research. Positivist research and interpretive research, therefore, can share modus tollens as a common scientific basis.

A common scientific basis is significant for two reasons. First, the availability of a common scientific basis – on which positivist research can build its theories with the help of, for example, mathematics and experimental design, and on which interpretive research can build its theories with the help of, for example, hermeneutics and the phenomenological reduction – weighs in favor of the argument that these two forms of research are not opposed and irreconcilable, but compatible (cf. Klein 2005; Lee 1991; Walsham 1995; Weber 2004). Second, recognition of a common scientific basis can set the stage for greater collaboration and mutual understanding between positivist researchers and interpretive researchers in the information-systems discipline, and thereby contribute to the creation of a cumulative body of information-systems knowledge. While different information-systems researchers use different methods, their acknowledgment of building on a common scientific basis can encourage them to see one another as members of the same scientific enterprise who are working together to reach the shared goal of advancing knowledge in the information-systems discipline.

Modus tollens, it must be emphasized, is not science itself, but instead provides a basis on which a structure of science can be built. Positivist research, interpretive research, action research, and design research can, and do, build upon this basis in their respective ways.

The Second Issue: the Fallacy of Affirming the Consequent

Another issue emerging from the MPMT framework is that researchers in information systems sometimes commit the "fallacy of affirming the consequent" in their reasoning. Familiarity with certain fundamental principles of logic would allow this erroneous reasoning to be identified and avoided.

Consider the three forms of reasoning in Figure 1. The third form of reasoning (in the last column, "if *p* is true, then *q* is true," "*q* is true," "therefore *p* is true") is neither modus ponens nor modus tollens. The flaw in such reasoning can be seen in this application of it: "If X is a human, then X is mortal," "Socrates is mortal," "therefore Socrates is a human." This reasoning is flawed because "Socrates" could designate a living being other than a human. This form of reasoning is what formal logic calls "the fallacy of affirming the consequent." It is appropriately called a fallacy because its conclusion does not follow from its major and minor premises. An ironic and counter-intuitive lesson following from this is that any evidence which agrees with a theory's prediction – where the theory is *p* and the prediction is *q* – may neither justify *nor even contribute to* the conclusion that the theory is true, no matter how much such evidence is produced. At best, it would allow the possibility that the theory is true to remain open.

The logic used in scientific reasoning, therefore, may not affirm or accept a theory as true. Fortunately, there is another option open to the logic of scientific reasoning. It is the option to deny or reject a theory as true. The logic of showing a theory to be wrong is the logic of modus tollens: “if the theory p is true, then its prediction q is true”; “the prediction q is not true”; “therefore, the theory p is false.” In Figure 1’s second column, this appears as “$p \supset q$,” “$\sim q$,” “$\therefore \sim p$.” The minor premise’s being $\sim q$ (i.e., the prediction is not true) is logically sufficient to justify the conclusion that the theory p is false, hence calling for the theory to be improved or replaced.

A place where the fallacy of affirming the consequent appears prominently in information-systems research is in its commentaries on the generalizability of case studies. Among information-systems researchers, the belief that the study of a single case is undesirable or deficient is widespread, even among those who are case researchers themselves; Lee and Baskerville provide numerous examples of this (2003, p. 223). However, requiring case study research to involve multiple sites or multiple cases for the sake of substantiating a theory is mistaken because this requirement presumes the (incorrect) logic in Figure 2, where the symbol “$\bullet$” means “and” and q_i is a prediction made by theory p in case study i. The major premise states: “if theory p is true, then its prediction q_1 in case study 1 will be true, and its prediction q_2 in case study 2 will be true, and its prediction q_3 in case study 3 will be true … and its prediction q_n in case study n will be true.” The minor premise states that all n predictions turn out to be true. The conclusion, in this illustration, is that the theory p is therefore true. This reasoning is incorrect because it commits the fallacy of affirming the consequent – and it commits the fallacy of affirming the consequent not just once, but n times.

No theory, no matter how great the number of its predictions or observational consequences[13] that turn out to be true, may ever be proven true – *not even once* – whether the theory is tested in a laboratory experiment, natural experiment, field experiment, or statistical experiment (no matter how large the sample size). The unprovability of theories is a

---

[13] The concept of “prediction” is a limiting case of the more general concept of “observational consequences.” See note e in the Notes for Table 2 and note h in the Notes for Tables 5 and 6.

fundamental principle of scientific reasoning. Evidence contradicting a theory is logically sufficient to show it to be false (where this would involve applying modus tollens), but no amount of evidence consistent with a theory may ever prove it to be true (lest the fallacy of affirming the consequent be committed). Familiarity with certain fundamental principles of logic and scientific reasoning in particular (the logic of modus ponens and the logic of modus tollens as used in the empirical testing of a theory) would allow the fallacy of affirming the consequent to be identified and avoided.

Another way of characterizing the unprovability of theories is that a successful result in empirical testing (i.e., turning up evidence that is consistent with a theory's prediction or observational consequences, especially if contradictory evidence was purposely being sought out) may, at best, only allow the theory to claim the provisional status of "good enough for now" or "has passed muster so far." A theory's status as valid is always provisional because there always remains the possibility for contradictory evidence to materialize in a future empirical test in a new setting. Argyris and Schön, like Popper, reserve the terms "corroborated" and "confirmed" to describe a theory which has achieved this provisional status.

Modus tollens provides a stronger basis for supporting scientific research when it is reinforced by explicit knowledge of, and vigilance against, the fallacy of affirming the consequent. Sound reasoning requires that the fallacy of affirming the consequent not be committed in any research that is to be considered rigorous.

The Third Issue: Summative Validity

Having expressed the logic of empirical inquiry in science in formal notation (Appendices A and B), including mathematical notation (Tables 4 and 6), we proceeded to search for published examples of it. For statistical behavioral research in information systems, which constitutes a large portion of positivist information-systems research, we sought out examples of modus tollens' major premise, minor premise, and conclusion as presented in Table

4. We encountered a surprise: We could find no such examples. And for interpretive research, we could find, arguably at best, just one or two examples of modus tollens as presented in Table 6. Because the reasoning used by information-systems researchers derives from, overlaps with, and contributes to the reasoning used by researchers in the other business-school disciplines, the social sciences, and the design sciences, one may speculate that the lack of such examples is not limited to information-systems research. The lack of examples has led us to distinguish two types of validity, formative and summative, where the use of modus tollens is required in order to establish the latter. As we will explain, we have come to hold the view that information-systems research has done much to establish the formative validity of its theories, but has rarely applied the logic of modus tollens to establish, in addition, their summative validity.

Differentiating Formative Validity and Summative Validity

We define formative validity as referring to the *process* by which a theory is *formed* or built (we will use the two terms "to form" and "to build" synonymously). We define summative validity as referring not to the theory-forming process, but to the *sum result* or *product* of the process, namely, the theory. A theory achieves formative validity by following one or another accepted procedure in the process of its being formed. A theory, once formed, achieves summative validity by surviving an empirical test that uses the logic of modus tollens (as shown in Table 4 for positivist theories tested through multivariate analysis and in Table 6 for interpretive theories).

For instance, for a theory to have formative validity in grounded-theory research, the theory's variables or constructs must emerge from, or be "grounded" in, the data rather than be taken entirely from a previously published theory and imposed on the current set of data. For a theory to have formative validity in statistical research, the process of building it must involve, among other things, data obtained through random or representative, rather than biased, sampling; thus a theory that is formulated from data describing a population's demographics, where the data were collected through biased sampling, would lack formative validity. Both of these examples focus on the quality of the theory-forming *process*, rather than on the quality of the theory itself as the *sum result* or *product* of the process. In contrast, for a theory to have summative validity, the theory must survive empirical test involving modus tollens, regardless of whatever way in which it was built, where survival is indicated by the observation of evidence consistent with what the theory predicts. Even if the theory was built in an accepted way so as to achieve formative validity, the theory could still fail an empirical test and thereby lack summative validity.

Some additional examples of summative validity and formative validity, outside the realm of science, can help to clarify how they are different. The teachers in a particular school may faithfully and properly teach all the required topics in the required ways to their students, but the students could still fail to learn what the teachers taught; the educational process in this school could claim formative validity, but the result or product of this process would lack summative validity.[14] As a systems developer, one may properly follow all the steps in an accepted systems development methodology when building a new information system (hence achieving formative validity), but still end up with an information system that is a failure (therefore indicating the lack of summative validity). Finally, we may properly follow all the steps in the recipe for baking a cake (hence indicating formative validity), but the resulting cake could still taste bad (therefore indicating no summative validity).

Of the two types, we consider summative validity to be more important. First, summative validity – which is established by empirical tests whose results include no evidence refuting, but only evidence consistent with, the theory being tested – is a necessary feature of any theory to be considered rigorous. Second, even though a theory which has formative validity is

---

[14] This example is particularly appropriate because this paper derives its concepts of formative validity and summative validity from the field of education's concepts of formative assessment and summative assessment.

arguably more likely also to have summative validity because it was built systematically, summative validity does not necessarily follow. Consider our earlier examples of a qualitative researcher who correctly employs the coding procedures of grounded-theory research and a quantitative researcher who conducts sampling that is random rather than biased. In each case, where the researcher continues to form the theory in a way that is accepted and considered valid, one might rightfully argue that the resulting theory is more likely to turn out to be accurate and, therefore, also to have summative validity. However, the resulting product (an information system, a cake, a student's mastery of a topic, a theory), once formed, remains open for testing. In other words, formative validity is not a sufficient condition for summative validity. Furthermore, there is also the consideration that even if a process of forming a theory relies on intuition and serendipity rather than accepted procedures (so that the theory would lack formative validity), this process can nonetheless result in a theory with which the evidence is consistent in subsequent testing.[15]

A description of how statistical behavioral research in information systems has overlooked, but can still achieve, summative validity requires an extensive examination of certain aspects of statistical reasoning, which we cover in the next section. The extensiveness of the examination is justified by the dominating role that statistical behavioral research has had in the information-systems discipline. Our examination pertaining to interpretive research follows in the section after the next.

---

[15] Auyang (2006) provides interesting examples. One example is the famous case of August Kelulé's conceptualization of the benzene ring, in which carbon atoms are bonded to one another so as to form the shape of a hexagon. He conceptualized it when dreaming of "a snake biting its tail." Formulating or building a theory through dreaming or other forms of serendipity is distinct from empirically testing the theory once it has been formed. As long as the theory survives empirical testing, its origin makes no difference.

Where Positivist Research Needs to Establish Summative Validity

The MPMT framework sheds light on situations where positivist research in information systems needs to complete, or “follow through” to, the task of establishing summative validity. The focus here will be on positivist, behavioral information-systems research that employs sampling-based, multivariate, hypothesis-testing methods. Such research has excelled at identifying and measuring relationships theorized to exist between variables, where these relationships can then be plausibly included as components in a theory. Formative validity is achieved by identifying these relationships in a systematic way for the given population, where this involves, among other things, using the well known and widely accepted procedures for examining the statistical significance of the numerical values estimated for the coefficients of the independent variables in multivariate analyses. However, in practice, behavioral information-systems research that relies on multivariate analysis and statistical inference has not followed through to the point of testing the overall theory once formed, thereby leaving the overall theory untested and its summative validity unestablished.

Our description of how statistical behavioral research has not, but can, follow through to the point of establishing summative validity will cover four items: 1) aspects of current statistical practice, 2) the additional statistical analysis required for summative validity, 3) illustrations with three examples of positivist studies, and 4) replication and holdback samples.

Aspects of Current Statistical Practice: Consider, as part of an illustration of the testing that still needs to take place, a theory that is mathematically operationalized as “$Y=ß_0+ß_1X_1+ß_2X_2+\dots+ß_nX_n$.” (We can alternatively use the general form “$Y_i=f(X_{i1}, X_{i2}, \dots ,X_{in})$ for $i=1,2,\dots,m$,” but the simpler form will suffice for purposes of illustration.) The logic of modus tollens, as expressed mathematically in Table 4, will lead us to scrutinize something familiar which we routinely do, but do not question: the statistical hypothesis testing that concerns a coefficient $ß_i$ (where the null hypothesis is $ß_i=0$). The scrutiny is needed to bring out the difference between the statistical hypothesis testing regarding a coefficient $ß_i$ and the statistical hypothesis testing regarding the overall theory in its mathematically operationalized form ("$Y=ß_0+ß_1X_1+ß_2X_2+\ldots+ß_nX_n$"), of which each $ß_i$ is only a part. The coefficients $ß_i$ will receive a detailed discussion (particularly in the following subsection, with the heading, "The Additional Statistical Analysis Required for Summative Validity") so that we can better explain the difference between testing coefficients and testing the overall theory.

For an example to use in our scrutiny, we turn to the technology acceptance model (Davis, Bagozzi, and Warshaw 1989; also see Table 1), which provides a familiar example of "$Y=ß_0+ß_1X_1+ß_2X_2+_{\ldots}+ß_nX_n$," where Y refers to a person's "behavioral intention to use" a given technology, X_1 refers to this person's "attitude toward using" the technology, and X_2 refers to the same person's "perceived usefulness" of the technology. If we substitute BI, A, and U for Y, X_1, and X_2 respectively, we would then have the relation "$BI=ß_0+ß_1A+ß_2U$," which is one of the three familiar equations used in operationalizing the technology acceptance model.

The way in which statistical analysis typically unfolds includes the following two tasks. In one, the researcher takes a sample from the population where the theory is being examined and then uses the sample to calculate a number b_i to use as an estimate of the true but unknown value of the coefficient $ß_i$. In the other task, the researcher conducts a test of the null hypothesis, "$ß_i=0$" (which is the hypothesis that no relationship exists between independent variable X_i and dependent variable Y) by determining if b_i is statistically significant (or, equivalently stated, deciding if $ß_i$, the unknown value of which b_i is only a sample-based approximation, is indeed different from zero). Figure 3 provides some details of statistical reasoning used for the purpose of testing the hypothesis "$ß_i=0$." Finding that b_i is statistically significant would lead to the decision to reject the null hypothesis "$ß_i=0$." The importance of such a decision is that it would allow the sample-based estimate b_i to be put forward as evidence consistent with the relationship theorized to exist between the independent variable X_i and the dependent variable Y. A process that establishes relationships in this manner would be a process that contributes to building a theory.

In such a theory-forming effort, formative validity is achieved by using traditional methods of statistical inference (such as testing the null hypothesis, "$ß_i=0$," as just described) to identify relationships among variables. The theory's summative validity, however, would still need to be established. This is because the task of using sample data to calculate each b_i and the task of determining each b_i's statistical significance both presume that the variables are related to each other in the way that the theory specifies in "$Y=ß_0+ß_1X_1+ß_2X_2+_{\dots}+ß_nX_n$," *but the truth of the theory is exactly what is in question in the first place*. As we will show, empirical testing of a null hypothesis about a coefficient $ß_i$ is different from empirical testing of the overall theory. We may properly describe the task of using a sample to estimate each coefficient and the task of determining each such estimate's statistical significance as preceding, but not substituting for, an empirical test of the overall theory to determine its summative validity. In its current practice, statistical behavioral information-systems research has not followed through to the point of testing the overall theory. A description of such a test follows.

The Additional Statistical Analysis Required for Summative Validity: Consider a theory mathematically operationalized in the form of an equation with just two independent variables, "$Y= ß_0+ß_1X_1+ß_2X_2$," and that, when instantiated in a given population, takes the form, "$Y_{predicted}=1.5X_1+0.3X_2$," where this would be the result of taking a random sample to estimate each coefficient $ß_i$. The following description of theory testing pertains to the general form, "$Y_i=f(X_{i1}, X_{i2}, \dots ,X_{in})$ for $i=1,2,\dots,m$," but "$Y= ß_0+ß_1X_1+ß_2X_2$" will suffice for purposes of illustration. In this context, synonyms for "to estimate" include "to measure" and "to calibrate." And if we were to continue with the technology-acceptance-model illustration, "$Y_{predicted}=1.5X_1+0.3X_2$" would become "$BI_{predicted}=1.5A+0.3U$."

Where the sample-based estimates b_1=1.5 and b_2=0.3 of the coefficients $ß_1$ and $ß_2$ are determined to be statistically significant, an empirical test of the overall theory could be conducted by calculating what "$Y_{predicted}$=1.5X_1+0.3X_2" provides as the dependent variable's predicted value, $y_{predicted}$, and then comparing the predicted value to the observed value, y_{new}. This would require observing an additional, out-of-sample data point – i.e., one that is apart from and in addition to the data points constituting the sample from which each b_i was computed. The new data point would need to be taken from the same population so that the same coefficient estimates (b_1=1.5 and b_2=0.3) can be used to calculate the prediction, $y_{predicted}$. In the earlier illustration where Y, X_1, and X_2 refer to the variables BI, A, and U in the technology acceptance model, the new data point would consist of the respective numerical values observed for an additional research subject's behavioral intention to use the technology, the same research subject's attitude toward using the technology, and the subject's perceived usefulness for the technology. Suppose, for the sake of illustration, that there are no other dependent variables than Y (or BI) and that the values observed for this out-of-sample research subject are y_{new}=1.9, $x_{1,new}$=1.2, and $x_{2,new}$=1.0.

For this new data point, the numerical values $x_{1,new}$=1.2 and $x_{2,new}$=1.0 would be plugged into the calibrated relation "$Y_{predicted}$=1.5X_1+0.3X_2," resulting in $y_{predicted}$=2.1 as the numerical value predicted for the dependent variable. (In the notation of Table 4, $y_{predicted}$ is calculated from $f_k(x_1, x_2, \dots, x_n)$). The predicted value of the dependent variable ($y_{predicted}$=2.1) would then be compared to its observed value (y_{new}=1.9).

2.1 and 1.9 are obviously different from each other, but this difference would not automatically allow the researcher to conclude that the prediction has failed. The reason is that the researcher would still need to account for the fact that "$y_{predicted}$=2.1," in being calculated from the relation "$Y_{predicted}$=1.5X_1+0.3X_2," is affected by the error associated with the coefficients, 1.5 and 0.3, which are only estimates based on a sample. The researcher can account for this by establishing an allowed margin of error (taking the form of an interval) around the value of $y_{\text{predicted}}$, within which the value of y_{new} could fall and still allow the researcher to have sufficient confidence to make the decision that the value of y_{new} is evidence consistent with the theory. The range of values in the allowed margin of error is a prediction interval. The proposition "$y_{\text{predicted}}=y_{\text{new}}$" is then rejected only if y_{new} falls outside the prediction interval around $y_{\text{predicted}}$. Standard statistics textbooks provide formulas for constructing prediction intervals for dependent variables (e.g., for the bivariate case, formulas are offered by Wonnacott & Wonnacott, 1984, p. 349, and Neter, Wasserman & Whitmore, 1988, p. 626). The concept of a prediction interval stands in contrast to the concept of a confidence interval; the former refers to a dependent variable's value, $y_{\text{predicted}}$, around which an interval is constructed, whereas the latter refers to an interval that is constructed around an estimate b_i of an independent variable's coefficient.

In the situation where the value of y_{new} is so far away from the value of $y_{\text{predicted}}$ that the former falls outside the prediction interval, the researcher could attribute this to either of two possibilities. In one possibility, the theory is true, so that the disparity between $y_{\text{predicted}}$ and the observed value y_{new} can be written off as due to the sampling error which, first, was earlier incurred when measuring each $ß_i$ with the estimate b_i and then, second, was introduced into the calculation of the value $y_{\text{predicted}}$. In the other possibility, the theory is not true, so that the disparity between y_{new} and $y_{\text{predicted}}$ is simply the result of the theory's being wrong (i.e., the independent and dependent variables are not, after all, related to one another as indicated in the theory's mathematical operationalization, "$Y=ß_0+ß_1X_1+ß_2X_2$"). The greater the disparity between y_{new} and $y_{\text{predicted}}$, the greater the confidence the researcher may have in making the decision to choose the latter possibility, which would mean rejecting the theory instead of attributing the disparity to sampling error. As Figure 3 shows, the logic of modus tollens is used in reaching the conclusion about whether or not to reject the theory.

A general procedure for using statistical inference to test a theory which is mathematically operationalized as "$Y_i=f(X_{i1}, X_{i2,} \ldots ,X_{in})$ for i=1,2,…,m" includes the following tasks: 1) The desired level of statistical significance (α, such as .05) or the corresponding desired level of confidence, $100(1-\alpha)\%$, is chosen. Suppose $\alpha=.05$ is chosen, so that we are dealing with 95% prediction intervals. 2) A sample is taken for the purposes of measuring each $ß_i$ with a numerical estimate b_i and then hypothesis testing is conducted to determine whether this estimate is statistically significant. Where we make the generous assumption (for the sake of illustration) that each b_i is statistically significant, the researcher performs the following three tasks: 3) An out-of-sample data point, from the same population whose sample was used in measuring each $ß_i$, is observed. 4) For this out-of-sample data point, the theory's prediction (the dependent variable's numerical value) $y_{\text{predicted}}$ is calculated and then compared to its observed value y_{new}. 5) If the observation y_{new} falls outside the 95% prediction interval or, equivalently, if $p<.05$, then the researcher makes the decision at the 95% confidence level that the prediction fails.

More than one failed prediction would better justify the decision to reject the theory. For 95% prediction intervals, the expected number of predictions that fail, when the theory is true, is 5 out of every 100. This means that if 100 out-of-sample points are taken in order to make and test 100 predictions, the result in which 4, 5, or 6 of them fail could be considered "in the ball park," or consistent with the theory's being true. But what is the number of failed predictions which would be needed to justify a decision to reject the theory? We use the binomial distribution to determine the required number of failed predictions (see Appendix C). If 100 out-of-sample points are taken, then 12 (or more) failed predictions would be sufficient, at the .01 or 1% level of statistical significance, to justify a decision to reject the theory.

"Rejecting a theory," it should be noted, can call for replacing the *entire* theory with a completely new theory or a rival theory, but "rejecting a theory" can also mean something much less drastic: It can call for rejecting *just the current formulation* of the theory, which can involve making some adjustments in it (e.g., retaining most or all of the independent variables but changing the relationships between them, adding a new independent variable, introducing a moderating relationship, etc.).

A word of caution is in order regarding the outcome in which 99 out of 100, or 999 out of 1000, data points fall inside their respective 95% prediction intervals. A larger number of successful predictions, while constituting more evidence consistent with the theory, would not be evidence that somehow better proves (than a smaller number of successful predictions) that the theory is true; to conclude otherwise would be to commit the fallacy of affirming the consequent. There is still a benefit, of course, associated with a larger number of out-of-sample data points that fall inside their respective prediction intervals. A larger number of successful predictions would increase the researcher's confidence (a subjective human sentiment) that the theory is true but, we must emphasize, would not somehow translate into greater validity (an attribute of the theory). We are using the term "confidence" here with the same meaning it has in the term "confidence interval."

Statistical techniques that use out-of-sample data points are nothing new. They are known as "cross validation" techniques, which include K-fold, leave-one-out, jackknife, and delete-d (Good, 1999, pp. 180-181). In general, the purpose of cross-validation techniques is to assess how well an estimated multivariate model fits the sample data from which it was estimated. In making such an assessment, a researcher sets aside part of the original sample for later use as out-of-sample data points, and uses the remaining data points as the sample for making estimates of population parameters (such as the coefficients $ß_i$) and thereby fitting the multivariate model to the data. For each one of the out-of-sample data points, the researcher can calculate the dependent variable's predicted value ($y_{\text{predicted}}$), compare it to the dependent variable's observed value (y_{new}), identify the difference between them ($y_{\text{predicted}} - y_{\text{new}}$) as an error, and then calculate a summary measure of all the errors across all the out-of-sample data points (such as the mean of the absolute values of all the errors).

What we offer that is new is the recognition that the use of a prediction interval (or the equivalent procedure of using a p-value) to test a theory's prediction ($y_{\text{predicted}}$) with an out-of-sample data point is tantamount to an experiment which can help to establish the theory's summative validity. It is the mathematical notation in which the logic of modus tollens is expressed in Table 4 that leads us to see that this manner of using of a prediction interval with an out-of-sample data point constitutes an experiment. Based on the results of such experimentation (which would involve multiple experiments, testing multiple predictions), the researcher may or may not justify the decision to reject the theory. This manner of testing a theory in order to establish its summative validity is further discussed below and in Appendix C.

Illustrations with Three Examples of Positivist Research: Table 1 provides three examples of positivist research conducted by well-known researchers and published in prominent journals: Davis, Bagozzi, and Warshaw (1989) in *Management Science*, Ang and Straub (1998) in *MIS Quarterly*, and Zhu and Kraemer (2005) in *Information Systems Research*. We are not randomly sampling these research studies for the purpose of somehow generalizing from them. Rather, we are specifically identifying them on the basis of their being conducted by well known scholars, appearing in major journals, and spanning a long period of time for the purpose of suggesting the magnitude of the implications that the issue of summative validity has for statistical behavioral research in information systems.

Each of the three studies examines a theory that its authors express in the form of a boxes-and-arrows diagram (or ellipses-and-arrows diagram). In Table 1, we re-express each theory by mathematically operationalizing it in form, "Y_i=f(X_{i1}, $X_{i2,}$ … ,X_{in}) for i=1,2,…,m." The authors of each study carry out, among other things, hypothesis testing associated with each sample-based estimate b_i of the coefficient $ß_i$, so that they demonstrate the formative validity of their respective theories. However, in none of the studies do the authors use any out-of-sample data points; thus, there is no instance in any of the studies in which the authors compare a predicted value of the dependent variable to its observed value. As a result, all three theories remain untested and their summative validity, not yet established. In this regard, these studies are typical of statistical behavioral research in information systems.

The MPMT framework can be readily applied in an effort to establish each theory's summative validity. Consider Davis *et al.*'s technology-acceptance model and its mathematical operationalization in the form of the relation, "$BI=ß_1A+ ß_2EOU$" (see Table 1). (For the sake of illustration, we will suppose that the technology-acceptance model involves no other relations.) The population that Davis *et al.* use for the purpose of statistical inference consists of all MBA students at the University of Michigan at the time of the study. In calibrating or fitting "$BI=ß_1A+ ß_2EOU$"[16] to this population, they take a sample (of MBA students) from it, resulting in "$BI_{predicted}=0.27A+0.48EOU$" where the two coefficients are statistically significant at the 0.01 and 0.001 levels, respectively (p. 992). Then, if Davis *et al.* were to apply the MPMT framework, they would proceed to consider an out-of-sample MBA student, for whom the numerical values of BI, A, and EOU would be observed (we designate them as bi_{new}, a_{new}, and eou_{new}). Suppose that, for this new research subject, the dependent variable BI's observed value bi_{new} were to fall outside the 95% prediction interval around the dependent variable's predicted value $bi_{predicted}$, thereby allowing the researcher to justify making a decision, at the 95% level of confidence, that the prediction fails. And suppose that the researcher tests a total of 100 such predictions, 12 of which fail. Then according to the second table in Appendix C, the researcher could justify making the decision, at the .01 or 1% level of significance, to reject the theory (here, the technology acceptance model) which made the 100 predictions, where the theory

---

[16] $ß_0=0$ because Davis *et al.* use standardized regression coefficients.

includes not only its mathematical operationalization, but also its verbally expressed assumptions, conditions, definitions, and other elements.

In Table 1, the proposition provided by Ang and Straub and the five propositions offered by Zhu and Kraemer would need to be similarly tested if the summative validity of their theories is to be established.

Past positivist research in information systems, in the form of statistical behavioral research, provides a wealth of theories whose formative validity has been established. Information-systems scholars who have engaged in statistical behavioral research deserve recognition for having met the Herculean challenge of providing, to the information-systems discipline, the needed theories in the first place. Should these studies have also established the summative validity of the theories they put forward? In general, there is no methodological requirement that a theory's summative validity must be established by the same researcher in the same study that establishes the theory's formative validity. Its summative validity can be established by the same or different researcher in a new, separate study. Research that offers new theories and establishes their formative validity will continue to be no less important and to make no less of a contribution than past research which has done this, such as the research by Ang and Straub (1998), Davis, Bagozzi, and Warshaw (1989), and Zhu and Kraemer (2005).

Replication and Holdback Samples: There are two additional concerns, associated with statistical behavioral research, which require attention. The first concern is: May the task of observing new, out-of-sample data points for the purpose of testing a theory be considered nothing more than a replication? The answer depends on what is meant by "replication."

Suppose a researcher uses sample data to compute each b_i in "$Y=ß_0+ß_1X_1+ß_2X_2+_{\ldots}+ß_nX_n$" and determines each b_i to be statistically significant. Then, in a second study, the researcher uses sample data from the *same* population to re-compute each b_i and re-determine each b_i to be statistically significant. The second study may be properly considered a replication of estimating the coefficients in "$Y=ß_0+ß_1X_1+ß_2X_2+_{\ldots}+ß_nX_n$" and a replication of determining their statistical significance, but this can occur without the overall theory's ever having been tested. If no out-of-sample data points are used in either or both studies, then there is no replication of any test of the theory. A study that does test the overall theory, which would call for using one or more additional out-of-sample data points from the same population that the sample was taken from, would be proceeding beyond the tasks of (re)estimating the coefficients and (re)establishing their statistical significance.

Suppose instead, in the second study, the researcher uses sample data not from the same, but a different population to re-compute each b_i and again determines each b_i to be statistically significant. The second study may be properly considered a replication of establishing the significance of the coefficients in "$Y_i=f(X_{i1}, X_{i2,} \ldots ,X_{in})$ for i=1,2,…,m," but again, re-determining each b_i to be statistically significant can occur without any empirical test of the overall theory ever having been conducted. If no out-of-sample data points are used in the second study (or, for that matter, the first study), then there is no replication of any test of the theory. Figure 3 provides a detailed comparison of the differences between testing coefficients in a theory and testing an overall theory.

To clarify this discussion, consider the study by Venkatesh, Morris, Davis, and Davis (2003) in which they examine eight theories (including the theory of reasoned action, the technology acceptance model, and the theory of planned behavior) in two different populations.[17] They estimate the coefficients $ß_i$ in each theory for each of the two populations, but they do not

---

[17] Venkatesh *et al.* observe two organizations where the use of technology is voluntary and two other organizations where its use is mandatory. Venkatesh *et al.* create two data sets. One data set is the result of pooling the sample data from the two voluntary-use organizations, where the data are treated as a sample from a single population. The same is done with the sample data from the two mandatory-use organizations. By conducting their statistical analysis in this way, Venkatesh *et al.* are examining the eight theories in one population and, again, in one other population.

use out-of-sample data points for testing the overall theories, thereby leaving each overall theory's summative validity unestablished. For each of the eight theories, Venkatesh *et al.* therefore can be described as performing, in one population, the task of estimating the coefficients and the task determining the estimates' significance and then, in a different population, replicating the two tasks; they do not, however, conduct any out-of-sample testing. Venkatesh *et al.* also examine a ninth theory (the unified theory of acceptance and use of technology, or UTAUT) in a new population (different from the first two),[18] but once again, no out-of-sample data points are used. We emphasize, as mentioned earlier, that a study which establishes a theory's formative validity (as Venkatesh *et al.* have done for UTAUT) need not also establish the theory's summative validity; the latter can be established in a future study that may be conducted by either Venkatesh *et al.* or other scholars.

The second concern, which is related to the first, is that a researcher who uses a holdback sample to replicate the estimation of the coefficients would not be testing the theory, but would just be accomplishing a re-estimation of the coefficients. Still, each data point in the holdback sample would have the potential to be used as an out-of-sample data point in an empirical test of the overall theory, where the researcher uses prediction intervals (or the corresponding procedure of testing with p-values) to compare the observed value of the dependent variable to its predicted value. In this manner, the number of such data points in the holdback sample can be the number of different experiments that the theory can be subjected to, where each such empirical test could be properly considered a replication of every other empirical test. In this regard, however, the term "sample" in "holdback sample" could be confusing; first, the term "sample" connotes using a sample to estimate a true but unknown value whereas out-of-sample data points are not used

---

[18] Venkatesh *et al.* observe two additional organizations (i.e., in addition to the four just mentioned) and create one new data set, which is the result of pooling the sample data from these two organizations. In this way, Venkatesh *et al.* are examining the ninth theory in a single population.

for this purpose and, second, the term “sample,” when used to refer to *out-of*-sample data points (whether a holdback sample or otherwise), would be misleading.

Statistical behavioral research is not alone in still needing to establish the summative validity of its theories. In this regard, it sits in the good company of interpretive research.

The Need to Establish Summative Validity in Interpretive Research

As Tables 5 and 6 indicate, the MPMT framework indicates that interpretive research, no less than positivist research, also needs to establish the summative validity of its theories in order for it to be considered rigorous. Interpretive research, however, has rarely completed, or followed through to, the task of establishing summative validity. A telling artifact is the *MIS Quarterly* special issue on intensive research. A requirement for a paper to be published in the special issue was for the paper to specify and apply a set of criteria with which one could substantiate the quality of the paper’s qualitative research. The first five studies in Table 7 involved interpretive research.

Table 7 shows the different validity criteria offered by the six papers. In this essay, we are categorizing almost all the criteria as falling under the heading of formative validity. For the sake of argument, suppose that some of the criteria that we have classified under formative validity were to be reclassified under the heading of summative validity. Even so, none of the papers differentiates summative validity from formative validity and, therefore, none of them recognizes or treats the former as a necessary condition for a theory to achieve in order to be considered scientific.

In applying the earlier metaphor of baking-*versus*-tasting a cake, we regard most of the criteria in Table 7 as more akin to the ingredients that go into baking a cake (where formative validity would be concerned with whether the right ingredients are used) than to the quality of the resulting cake (where summative validity would be concerned with testing whether the resulting cake tastes good). In Table 7, examples of ingredients include: “The Principle of

Contextualization," "Particularizing everyday life," "Self Revealing Writing," "providing rich detail about the organization," "Selection of appropriate field setting and methods," and "How do we capture the relevant knowledge?" These examples refer to research activities involved in forming a theory, not those involved in assessing the validity of a theory after it has been formed. Such an assessment, to determine if the evidence is consistent with the theory's observational consequences, would involve empirical testing using the logic of modus tollens.

Interpretive research, like positivist research, already has ways to establish the summative validity of a theory which are already available and ready for researchers to apply. For example, Sanday (1983) offers the following: "If, after having completed [the development of a theory by doing an] ethnography, the observer can communicate the rules for proper and predictable conduct as judged by the people studied, he or she has produced a successful product. The ethnographer is like the linguist who has recorded a foreign language so that others can learn the rules for producing intelligible speech in that language. As Frake (quoted in Wolcott, 1975: 121) says, the adequacy of ethnography is to be evaluated 'by the ability of a stranger to the culture (who may be the ethnographer) to use the ethnography's statements as instructions for appropriately anticipating the scenes of the society.' " In the major premise of modus tollens, "if *p* is true, then *q* is true," *p* could be "the researcher's ethnographically developed theory of the culture and social structure of the people in organization A" and *q* could be "after the researcher communicates her theory of the rules for proper and predictable conduct in this society to a stranger S, the stranger S will subsequently be successful in anticipating, and not surprised by, the people's behaviors in the upcoming scenes C and D in the society."

Schutz (1962) also offers a way to assess the summative validity of an interpretive theory (p. 64): "[E]ach term in a [theory] of human action must be constructed in such a way that a human act performed within the real world by an individual actor as indicated by the [theory] would be understandable to the actor himself as well as to his fellow-men in terms of the common-sense interpretation of everyday life." Here, *p* could be "the theory of human action T" and *q* could be "the human acts Q and R, performed by the vice president for operations at Acme Corporation as indicated by the theory of human action T, will be understandable to the vice president herself as well as to her colleagues in terms of their own common-sense interpretation of everyday work life at Acme Corporation." The ways of using "*p*" and "*q*" for the purpose of theory testing as suggested by Sanday and by Schutz are instances of the general form of the logics appearing in Tables 5 and 6.

Past qualitative and interpretive research in information systems, like past quantitative and positivist research, provides a wealth of theories with formative validity and should therefore be credited for having moved the information-systems discipline forward by accomplishing the major task of providing needed theories in the first place. Just as for quantitative and positivist research, qualitative and interpretive research that offers new theories and establishes their formative validity will continue to be no less important and to make no less of a contribution than past research which has done this, such as the research appearing in the *MIS Quarterly* special issue on qualitative research.

Additional Implications of the Modus Ponens, Modus Tollens Framework

We identify three additional major implications of the MPMT framework: the need to test old theories for summative validity; rigor in relevant research; and new directions needed in methodological research.

The Need to Test Old Theories for Summative Validity

For the most part, we have not empirically tested our theories. We need to test our theories rather than just build them and fit them to the empirical settings where we formulate them. By adding the use of prediction intervals to existing cross-validation techniques, we can establish the summative validity of theories which have already been published (such as those in the studies by Ang and Straub, 1998; Davis et al., 1989; Venkatesh et al., 2003; and Zhu and

Kraemer, 2005) *without collecting any new data*. The basic idea is to set aside part of the original sample as out-of-sample data points. One way to do this is to take a sub-sample from the original sample used in the published study, to use the sub-sample to estimate the multivariate model, and then to use each one of the remaining data points in a different experiment testing the theory. Another way is to use the original sample, but to leave out just one data point, where the remaining data points are used to estimate the model and the left-out data point is used in an experiment to test the theory; furthermore, each data point takes a turn as the left-out-data point, thereby allowing the researcher to conduct as many experiments as there are data points in the sample. Appendix C illustrates the combined use of the latter cross validation procedure (known as "leave one out") with prediction intervals. Furthermore, just as statistical behavioral theories can be re-examined with the original data, the same can be done with qualitative and interpretive theories, especially if the data were systematically archived (for instance, using what Yin, 1999, has called a "case study database").

Rigor in Relevant Research

In this essay, we have focused on rigor in positivist research and interpretive research, but the MPMT framework also has implications for relevance. Two forms of research known for emphasizing relevance are action research (Baskerville and Myers 2004) and design research (Hevner, March, Park, and Ram 2004). The discipline of information systems is not a basic or pure science; it is a field of professional study with constituents in government, business, and other organizations to whom it has obligations to offer research of practical value. The MPMT framework can provide a scientific basis for rigor not only in positivist and interpretive research, but also in action research and design research. Just as the logics of modus ponens and of modus tollens are blind to whether their propositions happen to be used in positivist research or interpretive research, they are blind to whether their propositions happen to be used in basic/pure research or relevant/applied research.

For action research, what would *p* and *q* stand for in the three statements of modus tollens ("if *p* is true, then *q* is true," "*q* is not true," "therefore, *p* is not true")? As shown in Table 8, *p* can stand for the proposition, "a theory of action for solving a certain organizational problem is effective" and *q* can stand for the proposition, "a particular action A prescribed by the theory of action for the organizational problem's instantiation P in organization O will solve the problem P." And in design research, also mentioned in Table 8, *p* could be "a design theory for solving a certain organizational problem is effective" and *q* could be "a particular artifact A prescribed by the design theory for the organizational problem's instantiation P in organization O will solve the problem P."[19]

Providing us with material for an example of the logic of modus tollens in action research are Kohli and Kettinger (2004). For one instance of the major premise "if *p* is true, then *q* is true" in their action research, the theory of action *p* is (p. 371) "Proposition A: Greater information transparency through the use of a performance monitoring information system (providing valid measures of behaviors and outcomes) will lead to greater goal congruence between the principal (hospital) and the agents (physicians)" and *q*, referring to an action which follows from the theory of action *p* and which achieves the given goal or solves the given problem, is (p. 372) "the physicians [in Kohli and Kettinger's field site will] begin to use the data [from the clinical DSS] to examine their practice and ultimately adopt quality improvement and/or cost cutting clinical procedures…" In applying the logic of modus tollens, Kohli and Kettinger's action research satisfies a necessary condition for rigor. Also following from the logic of modus tollens is that an observation (or even many observations) of a successful action (i.e., where the minor premise is "*q* is true") may never prove the theory of action to be correct (i.e., may never lead to the conclusion "*p* is true"), but the observation of an unsuccessful action (i.e., where the minor premise is "not *q*) can be logically sufficient to disprove the theory.

---

[19] The apparent parallel between action research and design science has been examined elsewhere (Lee, 2007).

Hevner, March, Park, and Ram (2004) interpret, as an example of design research, the work of Markus, Majchrzak, and Gasser (2002), who examine the goal of achieving success in "the development of information systems to support emergent knowledge processes (EKPs)" (Hevner *et al.*, p. 95). Markus *et al.* create an artifact, TOP Modeler (an implemented software system), with which to achieve this goal. "The TOP Modeler supports a development process incorporating the six principles for developing systems to support EKPs [emergent knowledge processes]" (Hevner *et al.*, p. 96), where we frame these six principles as constituting *p* in the major premise of modus tollens, "if *p* is true, then *q* is true." Markus *et al.* refer to them as "EKP Design Theory Principles" (p. 188), which are "Design for Customer Engagement by Seeking Out Naïve Users" (p. 188) "Design for Knowledge Translation Through Radical Iteration with Functional Prototypes" (p. 193), "Design for *Offline* Action" (p. 196), "Integrate Expert Knowledge with Local Knowledge Sharing" (p. 196), "Design for Implicit Guidance Through a Dialectical Development Process" (p. 199), and "Componentize Everything, Including the Knowledge-base" (p. 202). If *p* is true (i.e., if the proposition – that the six design principles are valid – is true), then *q* is true (i.e., the proposition – that the TOP Modeler, a particular artifact implementing these design principles, will be successful in achieving the given goal – is true). In applying the logic of modus tollens, Markus *et al.*'s research satisfies a necessary condition for rigor. And following the same logic as for Kohli & Kettinger, one may never declare the proposition *p* – here, that the six design principles are valid – to be true, no matter how many artifacts produce results consistent with the design principles; at best, whenever the proposition survives an empirical test, one may declare it to be "valid," where one also needs to make explicit that the status of "valid" is always tentative and can never be conclusively established.

The preceding discussion shows that relevant research, such as action research and design research, are as suitable for application of certain principles of logic in general and scientific reasoning in particular as are quantitative, positivist research and qualitative, interpretive research. The rigor of science is therefore achievable in relevant research *not necessarily by adopting the procedures of sampling-based multivariate hypothesis testing or other quantitative methods*, but by adopting the MPMT framework. Relevant research, such as action research and design science, certainly may exercise the option of adopting the procedures of sampling-based multivariate hypothesis testing, but these procedures are never, by themselves, necessary or sufficient for action research, design research, or any other form of inquiry to achieve rigor.

New Directions Needed in Methodological Research

This line of reasoning leads to the third and last implication of the MPMT framework to be mentioned. There is a need to change to a new direction in the ongoing development of research methods in the information-systems discipline. Currently, in positivist information systems research, there is an emphasis on developing increasingly rigorous methodological techniques that address formative validity (i.e., increasingly rigorous ways of measuring $ß_i$ and testing for the statistical significance of its estimated value, b_i); however, there is no less of a need to develop rigorous techniques for empirically testing an overall theory so as to establish its summative validity. And as Table 7 suggests, interpretive information systems research has also concentrated on the development of research methods that address formative validity and therefore also needs to turn some attention to the task of developing research methods for addressing summative validity.

Conclusion

The basic argument that we have been making is actually an old one. There is nothing new about modus tollens or formal logic in general, but a call for the conscientious application of such logic in empirical inquiry *is* new. Our call for returning to the basics of reasoning is altogether compatible with current streams of positivist, interpretive, action, and design research. The MPMT framework and the notions of formative validity and summative validity affirm what these streams of research have already accomplished, as well as point to what remains to be done.

We may conclude that the information-systems discipline, in having built up a large store of theories that have formative validity, has plentiful material with which to build up an equally large store of theories that will have, in addition, summative validity and perhaps even relevance. How would we recognize the situation in which the discipline of information systems has come to accept the requisite role of modus tollens in establishing summative validity? The signs may include the following: 1) information-systems researchers who now separate themselves based on the approaches they take – which include positivist research, interpretive research, action research, and design research – will see themselves as members of the same team who, albeit playing different positions on the team, are pursuing the common goal of building a cumulative body of knowledge in the information-systems discipline, which they can do by building on a common scientific basis; 2) information-systems researchers will no longer be committing the fallacy of affirming the consequent, particularly in their commentaries on case studies; and 3) the different research streams in the information-systems discipline will be giving due recognition to summative validity, whether established by the same researcher in the same study that demonstrates the theory's formative validity or by the same or different researcher in a completely separate study. And studies that establish the formative validity of theories (such as the studies by Ang and Straub, 1998; Davis et al., 1989; Venkatesh et al., 2003; and Zhu and Kraemer, 2005) will continue to be no less important than they are now.

One may always choose to define a scientific basis so that, by definition, it would require additional elements, such as mathematical propositions, numerical data, and statistical analysis. Still, regardless of whatever additional elements one might favor, no credible scientific research may ignore the fundamentals of formal logic. The discipline of information systems needs to consider returning to the basics – in particular, the logic of modus ponens and the logic of modus tollens. And in doing so, the information-systems discipline will fare better in its efforts to achieve research that is rigorous and relevant.

REFERENCES

Ang, S., and Straub, D. W. “Production and Transaction Economies and IS Outsourcing: A Study of the U.S. Banking Industry,” *MIS Quarterly* (22:4), 1998, pp. 535-552.

Argyris, C. *On Organizational Learning*, Blackwell Publishing, Cambridge, Massachusetts, 1999.

Argyris, C. and Schön, D. A. Theory in Practice: *Increasing Professional Effectiveness*, Jossey Bass, San Francisco, 1974.

Auyang, S. Y. “Chance and the Prepared Mind in Drug Discovery,” http://www.creatingtechnology.org/biomed/chance.pdf (accessed 17 December 2006).

Baskerville, R., and Myers, M. D. “Special Issue on Action Research in Information Systems: Making IS Research Relevant to Practice – Foreword,” *MIS Quarterly* (28:3), 2004, pp. 329-336.

Benbasat, I., and Zmud, R. W. “Empirical Research in Information Systems: The Practice of Relevance,” *MIS Quarterly* (23:1), 1999, pp. 3-16.

Bernstein R. *The Restructuring of Social and Political Theory*, New York: Harcourt Brace Jovanovich, 1976.

Davis, F. D., Bagozzi, R. P. and Warshaw, P. R. “User Acceptance of Computer Technology: A Comparison of Two Theoretical Models,” *Management Science* (35:8), 1989, pp. 982-1003.

Føllesdal, D. “Hermeneutics and the Hypothetico-Deductive Method,” *Readings in the Philosophy of Social Science*, M. Martin and L. C. McIntyre (eds.), MIT Press, Cambridge, Massachusetts, 1994, pp. 233-245.

Good, P. *Resampling Methods*, Birkhäuser: Boston and Berlin, 1999.

Gopal, A., and Prasad, P. “Understanding GDSS in Symbolic Context: Shifting the Focus from Technology to Interaction,” *MIS Quarterly* (24:3), 2000, pp. 509-546.

Gregor, S. “The Nature of Theory in Information Systems,” *MIS Quarterly* (30:3), 2006, pp. 611-642.

Hevner, A. R., March, S. T., Park, J. and Ram, S. “Design Science in Information Systems Research,” *MIS Quarterly* (28:1), 2004, pp. 75-105.

Klein, H. K. “The Research Potential of Hermeneutic Field Studies,” Joint Ph.D. Workshop on Information Systems (Louisiana State University, University of Texas at San Antonio, University of Houston), E. J. Ourso College of Business Administration, Louisiana State University, 10 February 2005, http://projects.bus.lsu.edu/students/evgeny/PhD_workshop/Presentations/HermeneuticFS_v2SOM050208.ppt (accessed 17 December 2006).

Klein, H. K., and Myers, M. D. “A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems,” *MIS Quarterly* (23:1), 1999, pp. 67-94.

Klenk, V. *Understanding Symbolic Logic*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

Kohli, R., and Kettinger, W. J. “Informating the Clan: Controlling Physicians’ Costs and Outcomes,” *MIS Quarterly* (28:3), 2004, pp. 363-394.

Kolakowski, L. *The Alienation of Reason: A History of Positivist Thought*, Doubleday: Garden City, 1968.

Kuhn, T. S. *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, 1962.

Kuhn, T. S. “Second Thoughts on Paradigms,” in *The Essential Tension*, University of Chicago Press, 1977, pp. 293-319.

Lee, A. S. “Action is an Artifact: What Action Research and Design Science Offer to Each,” in *Information Systems Action Research: An Applied View of Emerging Concepts*, N. Koch (ed.), Springer US, New York, 2007, pp. 43-60.

Lee, A. S. “Researching MIS,” *Rethinking Management Information Systems: An Interdisciplinary Perspective*, W. Currie and B. Galliers (eds.), Oxford University Press, New York, 1999, pp. 7-27.

Lee, A. S. “Integrating Positivist and Interpretive Approaches to Organizational Research,” *Organization Science* (2:4), 1991, pp. 342-365.

Lee, A. S. “A Scientific Methodology for MIS Case Studies,” *MIS Quarterly* (13:1), 1989, pp. 33-50.

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

Markus, M. L. “The Qualitative Difference in Information Systems Research and Practice,” *Information Systems and Qualitative Research: Proceedings of the IFIP TC8 WG 8.2 Meeting in Philadelphia*, A. S. Lee and J. Liebenau (eds.), Chapman & Hall, Ltd., London, 1997, pp. 11-27.

Markus, M. L., Majchrzak, A. and Gasser, L. “A Design Theory for Systems that Support Emergent Knowledge Processes,” *MIS Quarterly* (26:3), 2002, pp. 179-212.

Mendenhall. W. and Sincich, T. *A Second Course in Statistics: Regression Analysis*, Prentice Hall: Upper Saddle River, New Jersey, 2003.

Nelson, K. M., Nadkarni, S., Narayanan, V. K., and Ghods, M. “Understanding Software Operations Support Expertise: A Revealed Causal Mapping Approach,” *MIS Quarterly* (24:3), 2000, pp. 475-507.

Neter, J., Wasserman, W. and Whitmore, G. A. *Applied Statistics*, Allyn and Bacon, Inc.: Boston and London, 1988.

Neyman, J. and Pearson, E.S. “On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I,” *Biometrika*, (Vol. 20A, No. 1/2), 1928, pp. 175-240.

Orlikowski, W. J. and Baroudi, J. J. “Studying Information Technology in Organizations: Research Approaches and Assumptions,” *Information Systems Research* (2:1), 1991, pp. 1-28.

Popper, K. R. *The Logic of Scientific Discovery*, Harper Torchbooks, New York, 1968a.

Popper, K. R. *Conjectures and Refutations: The Growth of Scientific Knowledge*, Harper Torchbooks, New York, 1968.

Ricoeur, P. *From Text to Action*, Northwestern University Press, Evanston, 1991.

Sanday, P. R. “The Ethnographic Paradigm(s),” *Qualitative Methodology,* John Van Maanen (ed.), Sage Publications, Beverly Hills, CA, 1983, pp. 19-36.

Sarker, S., and Lee, A. S. “Does the Use of Computer-Based BPC Tools Contribute to Redesign Effectiveness? Insights from a Hermeneutic Study,” *IEEE Transactions on Engineering Management* (53:1), 2006, pp. 130-145.

Schultze, U. “A Confessional Account of an Ethnography about Knowledge Work,” *MIS Quarterly* (24:1), 2000, pp. 3-41.

Schutz, A. “Concept and Theory Formation in the Social Sciences,” *Collected Papers I: The Problem of Social Reality,* M. Natanson (ed.), Kluwer Academic Publishers, Boston, 1962, pp. 48-66.

Trauth, E. M. and Jessup, L. M. “Understanding Computer-Mediated Discussions: Positivist and Interpretive Analyses of Group Support System Use,” *MIS Quarterly* (24:1), 2000, pp. 43-79.

Walsham, G. “The Emergence of Interpretivism in IS Research,” *Information Systems Research* (6:4), 1995, pp. 376-394.

Walsham, G., and Sahay, S. “GIS for District-Level Administration in India: Problems and Opportunities,” *MIS Quarterly* (23:1), 1999, pp. 39-66.

Weber, R. "The Rhetoric of Positivism versus Interpretivism: A Personal View," *MIS Quarterly* (28:1), 2004, pp. iii-xii.

Wonnacott, T. H. and Wonnacott, R. J. *Introductory Statistics for Business and Economics*, John Wiley & Sons, New York, 1984.

Yin, R. *Case Study Research*, Newbury Park, CA: Sage Publications, 1994.

Zhu, K., and Kraemer, K. L. "Post-Adoption Variations in Usage and Value of E-business by Organizations: Cross-Country Evidence from the Retail Industry," *Information Systems Research* (16:1), 2005, pp. 61-84.

	modus ponens (MP)	modus tollens (MT)	fallacy of affirming the consequent
major premise	$p \supset q$	$p \supset q$	$p \supset q$
minor premise	p	$\sim q$	q
conclusion	$\therefore q$	$\therefore \sim p$	$\therefore p$

Figure 1

major premise	$p \supset q_1 \bullet q_2 \bullet q_3 \bullet \ldots \bullet q_n$
minor premise	$q_1 \bullet q_2 \bullet q_3 \bullet \ldots \bullet q_n$
conclusion	$\therefore p$

Figure 2

Modus Tollens and Statistical Inference

Statements in the social sciences are often statistical or probabilistic in nature. Their statistical or probabilistic nature is readily captured by the logic of modus tollens. In statistical hypothesis testing that concerns the coefficient $ß_i$ or the proposition "$Y=ß_0+ß_1X_1+ß_2X_2+_{\ldots}+ß_nX_n$," modus tollens takes the form where the major premise is "if H_0 is true, then p-value>α," the minor premise is "p-value<α," and the conclusion is "therefore, reject H_0 as true." The use of modus tollens in statistical hypothesis testing appears in the work of Neyman and Pearson (1928) and the work Popper, who refers to it in his "methodological rule" (1998, page 191).

	the meaning of H_0	the meaning of the minor premise, "p-value<α"	the meaning of the minor premise, "p-value>α"
testing the null hypothesis regarding the coefficient $ß_i$	The null hypothesis H_0 is that $ß_i=0$.	b_i, which is a sample-based approximation of the true but unknown value of $ß_i$, falls *outside* the 100(1-α)% confidence interval around 0.	b_i, which is a sample-based approximation of the true but unknown value of $ß_i$, falls *inside* the 100(1-α)% confidence interval around 0.
testing the proposition "$Y=ß_0+ß_1X_1+ß_2X_2+_{\ldots}+ß_nX_n$"	The null hypothesis H_0 is that $Y=ß_0+ß_1X_1+ß_2X_2+_{\ldots}+ß_nX_n$. In null form, H_0 is that $Y-(ß_0+ß_1X_1+ß_2X_2+_{\ldots}+ß_nX_n)=0$.	The out-of-sample data point provides a value of y_{new} that falls *outside* the 100(1-α)% prediction interval around $y_{\text{predicted}}$, where $y_{\text{predicted}}$ is an approximation of the true but unknown value of the predicted value of the dependent value Y for this out-of-sample data point.	The out-of-sample data point provides a value of y_{new} that falls *inside* the 100(1-α)% prediction interval around $y_{\text{predicted}}$, where $y_{\text{predicted}}$ is an approximation of the true but unknown value of the predicted value of the dependent value Y for this out-of-sample data point.

Figure 3

Table 1 – three examples of positivist information-systems research with theories operationalized mathematically as “$Y_i=f(X_{i1}, X_{i2}, \dots ,X_{in})$ for i=1, 2, … , m”

	THEORY EXAMINED
Davis, F.D., Bagozzi, R.P. & Warshaw, P.R. “User Acceptance of Computer Technology: A Comparison of Two Theoretical Models,” *Management Science*, (35:8), 1989, pp. 982-1003.	The theory which Davis *et al.* examine is the technology acceptance model. They use a boxes-and-arrows diagram to depict the theory’s variables and relationships (p. 985), which corresponds to the following mathematical operationalization of the theory (p. 992): $BI = ß_{0,BI} + ß_{1,BI}A + ß_{2,BI}U$ $A = ß_{0,A} + ß_{1,A}U + ß_{2,A}EOU$ $U = ß_{0,U} + ß_{1,U}EOU$ This is an instance of the general form “$Y_i=f(X_{i1}, X_{i2}, \dots ,X_{in})$ for i=1, 2, … , m,” where a total of 3 independent variables (n=3) appear in 3 simultaneous equations and where each of the 3 dependent variables (m=3) requires its own equation. BI is behavioral intention; A is attitude toward using technology; U is perceived usefulness; and EOU is perceived ease of use.
Ang, S. & Straub, D.W. “Production and Transaction Economies and IS Outsourcing: A Study of the U.S. Banking Industry,” *MIS Quarterly*, (22:4), 1998, pp. 535-552.	Ang & Straub refer to their theory as “Research Model for Current Study” (see their Figure 1 on page 536). They use an ellipses-and-arrows diagram to depict the theory’s variables and relationships, which corresponds to the following mathematical operationalization of the theory: $ISO = ß_0 + ß_1PCA + ß_2TC + ß_3FSL + ß_4FSI$ This is an instance of the general form, “$Y_i=f(X_{i1}, X_{i2}, \dots ,X_{in})$ for i=1, 2, … , m,” where a total of 4 independent variables (n=4) and a single dependent variable (m=1) appear in just single equation (p. 992): ISO is the degree of information systems outsoucing; PCA, production cost advantage; TC, transaction cost; FSL, financial slack; and FSI, firm size.

Table 1 – three examples of positivist information-systems research with theories operationalized mathematically as “Y_i=f(X_{i1}, X_{i2}, … ,X_{in}) for i=1, 2, … , m”

	THEORY EXAMINED
Zhu, K. & Kraemer, K.L. “Post-Adoption Variations in Usage and Value of E-Business by Organizations: Cross-Country Evidence from the Retail Industry,” *Information Systems Research*, (16:1), 2005, pp. 61-84.	Zhu & Kraemer examine a theory, “an integrative model for e-business use and value” (p. 62). They use an ellipses-and-arrows diagram to depict the variables and relationships (p. 66 and p. 77), which corresponds to the following mathematical operationalization of the theory: IOS = ß$_{0,IOS}$ + ß$_{1,IOS}$EBV IIO = ß$_{0,IIO}$ + ß$_{1,IIO}$EBV IP = ß$_{0,IP}$ + ß$_{1,IP}$EBV EBV= ß$_{0,EBV}$ + ß$_{1,EBV}$FEF + ß$_{2,EBV}$EBU + ß$_{3,EBV}$BEI EBU= ß$_{0,EBU}$ + ß$_{1,EBU}$TC + ß$_{2,EBU}$S + ß$_{3,EBU}$IS + ß$_{4,EBU}$FC + ß$_{5,EBU}$CP + ß$_{6,EBU}$RS This is the mathematical form, “Y_i=f(X_{i1}, X_{i2}, … ,X_{in}) for i=1, 2, … , m,” where a total of 10 independent variables (n=10) appear in 5 simultaneous equations and where each of the 5 dependent variables (m=5) requires its own equation. IOS is impact on sales; IIO, impact on internal operation; IP, impact on procurement; EBV, e-business value; EBU, e-business use; FEF, front-end functionality; BEI, back-end integration; TC, technology competence; S, size; IS, international scope; FC, financial commitment; CP, competitive pressure; and RS, regulatory support.

Table 2 – the Logic of the Syllogism, Illustrated with Positivist Notation

the theory (mathematically operationalized) (major premise)	$Y = f(X_{1,} X_{2,} \dots , X_n)$
the initial conditions (minor premise)	$X_1 = x_1,\ X_2 = x_2,\ X_3 = x_3, \dots , X_n = x_n$
the prediction (conclusion)	$Y = f(x_1, x_2, x_3, \dots, x_n)$ or $Y = y_{\text{predicted}}$

Notes for Table 2

In positivist research that seeks to perform multivariate statistical testing, *the theory* can be depicted in the form of a boxes-and-arrows (or ellipses-and-arrows) diagram, which can be mathematically operationalized in the form, "$Y_i = f(X_{i1}, X_{i2,} \dots , X_{in})$ for i=1, 2, … , m," where n is the number of independent variables and m is the number of dependent variables or equations. The simplified case, "$Y = f(X_{1,} X_{2,} \dots , X_n)$," is sufficient for illustrative purposes.

a. "Y" is the dependent variable. "X_i" is the $i^{\underline{\text{th}}}$ independent variable. "f" is the function that relates the variables to one another. "x_i" is a numerical value taken by the variable X_i.

   "$f(x_1, x_2, \dots , x_n)$" is the numerical value computed for $y_{\text{predicted}}$. It results from plugging the number x_i into every independent variable X_i. The numerical value that is then calculated and *predicted* for Y, $y_{\text{predicted}}$, is defined as equal to $f(x_1, x_2, \dots , x_n)$. In contrast, y_{new} is the numerical value *observed* for y (see point c, below) in an empirical test of the theory.

   The notation "$Y=f(X_{1,} X_{2,} \dots , X_n)$" stands for how the variables are related to one another, for which some examples (where n=3) are "$Y=\text{ß}_0+\text{ß}_1X_1+\text{ß}_2X_2+\text{ß}_3X_3$," "$Y=(\text{ß}_1X_1/\text{ß}_3X_3)^2+\text{ß}_2X_2$," and "$Y=(\text{ß}_1X_1+\text{ß}_2X_2)(\text{ß}_1X_1-\text{ß}_3X_3)$." In these examples, ß_i refers to the true but unknown numerical value for the coefficient of independent variable X_i. When statistical inference is used, ß_i is typically estimated with a sample-based number, to which this essay refers as b_i.

b. *The initial conditions* are facts describing an instance of the phenomenon that the theory is offered to explain. In a statistical, laboratory, field, or natural experiment, the initial conditions describe the phenomenon *without* or *before* the application of the experimental treatment .

c. *The prediction* is an assertion by the theory about something that should be observed about the phenomenon *with* or *after* the application of the experimental treatment (in particular, the dependent variable's calculated and predicted numerical value, $y_{\text{predicted}}$, should equal the dependent variable's observed numerical value, $y_{\text{predicted}}$). For this reason, the term "expected subsequent conditions" would actually make better sense than the term "prediction," as well as nicely complement the term "initial conditions."

d. The terms "initial conditions" and "prediction" suggest a temporal sequence or longitudinal analysis, but the concepts behind these terms also appear in cross-sectional research. Consider statistical experiments. They may properly involve only data that are all collected at one point in time. In such an experiment, the initial conditions would refer to values taken by the variables in the situation *without* the application of the "statistical treatment," and the prediction would reflect the change in value of the dependent variable in the situation *with* the statistical treatment.

e. One may use the term "observational consequence" in place of "prediction" for two reasons: first, this term refers to what one expects to *observe* in the given setting as a *consequence* of the theory's being true and, second, this term neither specifies nor implies any time element. As such, "observational consequence" is more general than and subsumes "prediction" and is even sufficiently general to be used in descriptions of the logic of the syllogism in interpretive research (see point h in the Notes for Tables 5 and 6).

Table 3 – Modus Ponens (MP) – positivist research

$p \supset q$	If the theory T_k, mathematically operationalized as "$Y = f_k(X_1, X_2, \ldots , X_n)$," is true, then the prediction "$y_{\text{predicted}} = y_{\text{new}}$" is true.
p	The theory T_k, mathematically operationalized as $Y = f_k(X_1, X_2, \ldots , X_n)$, is true.
$\therefore q$	Therefore the prediction "$y_{\text{predicted}} = y_{\text{new}}$" is true.

Table 4 – Modus Tollens (MT) – positivist research

$p \supset q$	If the theory T_k, mathematically operationalized as "$Y = f_k(X_1, X_2, \ldots , X_n)$," is true, then the prediction "$y_{\text{predicted}} = y_{\text{new}}$" is true.
$\sim q$	The prediction "$y_{\text{predicted}} = y_{\text{new}}$" is not true.
$\therefore \sim p$	Therefore the theory T_k, mathematically operationalized as "$Y = f_k(X_1, X_2, \ldots , X_n)$," is not true.

Notes for Tables 3 and 4

a. The syllogism in Table 2 (reproduced below in statements 1, 2, and 3) leads to statement 8, which is the major premise that appears in Table 3 and also in Table 4.

1.	"$Y = f(X_1, X_2, \ldots , X_n)$" is true.	major premise
2.	"$X_1 = x_1$, $X_2 = x_2$, $X_3 = x_3, \ldots , X_n = x_n$" is true.	minor premise
3.	Therefore "$Y = f(x_1, x_2, x_3, \ldots, x_n)$" is true.	conclusion from applying 1 to 2
4.	$y_{\text{predicted}} \equiv f(x_1, x_2, x_3, \ldots, x_n)$	definition of $y_{\text{predicted}}$
5.	Therefore "$Y = y_{\text{predicted}}$" is true.	restatement of 3 using 4
6.	If "$Y = f(X_1, X_2, \ldots , X_n)$" is true, then "$Y = y_{\text{predicted}}$" is true.	conditional proof, 1-5
7.	If "$Y = y_{\text{predicted}}$" is true, then "$y_{\text{predicted}} = y_{\text{new}}$" is true.	premise
8.	Therefore if "$Y = f(X_1, X_2, \ldots , X_n)$" is true, then "$y_{\text{predicted}} = y_{\text{new}}$" is true.	hypothetical syllogism, 6-7.

b. p = {the theory T_k, mathematically operationalized as "$Y = f_k(X_{1,} X_{2,} \ldots , X_n)$" is true}. q = {the prediction, "$y_{\text{predicted}} = y_{\text{new}}$," is true}

c. The terms Y, y_{new}, $y_{\text{predicted}}$, X_i, x_i, f, and n have the same definitions as in the Notes for Table 2.

d. In T_k, T refers to the theory being tested in its mathematically operationalized form and k refers to the form taken by the theory in the $k^{\underline{th}}$ round of testing. In f_k, k refers to the form taken by the function in the $k^{\underline{th}}$ round of testing.

e. Tables 3 and 4 can be readily adjusted to cover the general case in which a theory takes the form, "$Y_i=f(X_{i1}, X_{i2,} \ldots ,X_{in})$ for i=1, 2, … , m." The less complicated form "$Y = f_k(X_{1,} X_{2,} \ldots , X_n)$" suffices for purposes of illustration.

f. An empirical disconfirmation in a properly designed and conducted test of the theory T_k mathematically operationalized as "$Y = f_k(X_{1,} X_{2,} \ldots , X_n)$" would necessitate that it be replaced by an improved version or a completely new version, either of which we would designate as T_{k+1}. Theory T_{k+1} would then need to undergo empirical testing and also face the possibility of disconfirmation. Its disconfirmation would necessitate that it be replaced by yet another improved or completely new version, T_{k+2}, which would then also need to undergo empirical testing, and so forth.

Table 5 – Modus Ponens (MP) – interpretive research

$p \supset q$	If a reader's interpretation I_k of a text is a valid hermeneutic interpretation, then the reader's interpretation of any passage P_i or set of passages (e.g., P_2, P_3, P_5) in the text will not give rise to a contradiction, inconsistency, or other anomaly with regard to the reader's interpretation of any other passage or set of passages in the text.
p	A reader's interpretation I_k of a text is a valid hermeneutic interpretation.
$\therefore q$	Therefore the reader's interpretation of any passage P_i or set of passages (e.g., P_2, P_3, P_5) in the text will not give rise to a contradiction, inconsistency, or other anomaly with regard to the reader's interpretation of any other passage or set of passages in the text.

Table 6 – Modus Tollens (MT) – interpretive research

$p \supset q$	If a reader's interpretation I_k of a text is a valid hermeneutic interpretation, then the reader's interpretation of any passage P_i or set of passages (e.g., P_2, P_3, P_5) in the text will not give rise to a contradiction, inconsistency, or other anomaly with regard to the reader's interpretation of any other passage or set of passages in the text.
$\sim q$	A reader's interpretation of a new passage in the text (a passage she did not read previously) gives rise to a contradiction, inconsistency, or other anomaly with regard to the interpretation she made of another passage or set of passages she previously read.
$\therefore \sim p$	Therefore the reader's interpretation I_k of the text is not a valid hermeneutic interpretation.

Notes for Tables 5 and 6

a. In I_k, I refers to a reader's interpretation of a text and k refers to the reader's $k^{\underline{th}}$ reading and $k^{\underline{th}}$ interpretation of the text. We define the term "interpretive theory" to be such an interpretation, I_k.

b. p = {a reader's interpretation I_k of a text is a valid hermeneutic interpretation}. q = {the reader's interpretation of any passage P_i or set of passages (e.g., P_2, P_3, P_5) in the text will not give rise to a contradiction, inconsistency, or other anomaly with regard to the reader's interpretation of any other passage or set of passages in the text}.

c. Constituting a text are the passages $P_1, P_2, \ldots, P_n$. The passages $P_1, P_2, \ldots, P_n$ may or may not be mutually exclusive, but are mutually exhaustive of the text.

d. In interpreting a text, a reader strives for an interpretation of each of the different passages making up the text (the "parts") as well as the overall text (the "whole").

e. The conclusion "therefore the reader's interpretation I_k of the text is not a valid hermeneutic interpretation," if reached, would call for I_k it to be replaced by an improved version or a completely new interpretation. This interpretation, I_{k+1}, would then be assessed as to whether it is a valid hermeneutic interpretation of the text, just as I_k was assessed. The process of crafting an interpretation in this iterative manner, involving the logic of modus tollens, is precisely the logic of the hermeneutic circle, which Klein and Myers (1999) characterize as the overarching principle of interpretive research.

f. Interpretive research may involve the interpretation of not only text, but also text analogues. For example, if a researcher's interpretation of the body of shared rules for allowable behavior in a particular organization is valid, then the researcher's observation of any action by an organizational member will be consistent with this interpretation. Here, the text analogues are "the body of shared rules for allowable behavior" in this organization and "any action by an organizational member."

g. A major point following from the MPMT framework is that even when an interpretation is assessed to be a valid interpretation, this assessment does *not* allow the reader to describe the interpretation as the "final" interpretation, the "definitive" interpretation, or "the" correct interpretation. There always remains the possibility for a different interpretation to be formulated so as also to be a valid interpretation of the same text or text analogue. Interpretive researchers regard as acceptable the co-existence of more than one interpretation of the same text or text analogue; interpretive researchers acknowledge the reality of multiple realities. Another reason for not describing any interpretation as final or definitive is that there is always the possibility for an inconsistency in the interpretation to arise when reading an additional passage (e.g., the next "paragraph" or next "chapter") of the text.

h. Positivist research and interpretive research have a difference in terminology with regard to *q* as it appears Tables 2, 3, 4, 5, and 6. The positivist term for *q*, "prediction," is rarely or never used in interpretive research. Fortunately, the philosophy of science already has a term, "observational consequence" (explained above in the Notes for Table 2), which is sufficiently general to subsume "prediction" and also to be appropriate for interpretive research. For instance, the prediction or *observational consequence* that follows from the positivist theory f=ma in a setting where the mass, m, is 10 and the acceleration, a, is 4 is that, if the theory is true, the force will be 40. As for interpretive research, consider the theory that e-mail can support rich communication only if the e-mail users share a socially constructed reality; the *observational consequence* of this theory for an actual organization where the e-mail users do not share a socially constructed reality is that, if the theory is true, the e-mail communication will not be rich. Also worth noting is that, in both the positivist and interpretive examples, the theory's observational consequence for the given setting would still need to be compared to an observation of what actually happens in that setting.

i. Scholars have already noted the parallels between positivist theory development and interpretive theory development. Sarker and Lee (2006, p. 134) refer to Smith (1993) who uses the positivist term "falsification" in describing the development of a hermeneutic interpretation. Sarker and Lee also refer to Ricoeur (1991, pp. 159–160) who eschews the term “verification” and instead applies the concept, associated with positivist theory, that the truth of a theory may never be verified: “As concerns the procedures of validation with which we test our guesses, I agree with Hirsch that they are a lot closer to a logic of probability than to a logic of empirical verification. To show that an interpretation is more probable in the light of what is known … is something other than showing that a conclusion is true…” Also, in “Hermeneutics and the Hypothetico-Deductive Method,” Føllesdal (1994) states (emphasis in the original, p. 233): “*the hermeneutic method is the hypothetico-deductive method applied to meaningful material (texts, works of art, actions, etc.).*”

Table 7 – formative criteria and summative criteria in the articles in the *MIS Quarterly* special issue on intensive research

	ARTICLE'S FORMATIVE CRITERIA	ARTICLE'S SUMMATIVE CRITERIA
Klein, H.K. & Myers, M.D. "A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems," (23:1), 1999, p. 72.	The Fundamental Principle of the Hermeneutic Circle The Principle of Contextualization The Principle of Interaction Between the Researchers and the Subjects The Principle of Abstraction and Generalization The Principle of Multiple Interpretations The Principle of Suspicion	The Principle of Dialogical Reasoning
Walsham, G. & Sahay, S. "GIS for District-Level Administration in India: Problems and Opportunities," (23:1), 1999, p. 59.	Particularizing everyday life Delineating authors' relationships in the field Depicting the disciplined pursuit and analysis of data Qualifying personal biases Normalizing unorthodox methodologies Legitimating the atypical Smoothing the contextable Dramatic anticipation Carving out room to reflect Provoking the recognition and examination of differences Imagining new possibilities	

Table 7 – formative criteria and summative criteria in the articles in the *MIS Quarterly* special issue on intensive research

	ARTICLE'S FORMATIVE CRITERIA	ARTICLE'S SUMMATIVE CRITERIA
Schultze, U. "A Confessional Account of an Ethnography About Knowledge Work," (24:1), 2000, p. 30.	*Authenticity (demonstrate that the ethnographic researcher was indeed immersed in the field)* *Plausibility (present the findings as relevant to the common concerns of the audience)* *Criticality (move readers to reexamine their own taken-for-granted assumptions)* *Self Revealing Writing* *Interlacing "actual" and confessional content*	
Trauth, E.M. & Jessup, L.M. "Understanding Computer-Mediated Discussions: Positivist and Interpretive Analysis of Group Support System Use," (24: 1), 2000, pp. 66-70.	**Triangulation**—the use of multiple sources, methods and investigators to provide corroborating evidence… **Authenticity**...In this research, authenticity refers to the interpretive rendering of both the discussion transcripts and the context from which they arose. By providing rich detail about the organization, the participants, the relevant perceptions, actions, and events, and the relevant issues, and by describing the local and broader contexts within which the research took place, we helped the reader to better sense the meaning of this context.	**Breakdown Resolution or Hermeneutic Circle**… we accomplished this through the resolution of breakdowns, to use the words of ethnography (Agar 1986), or through the hermeneutic circle, to use the words of hermeneutics. Both terms characterize interpretation as an iterative process of examining the particular in relation to the greater whole and revising meanings as these iterations progress. When an anomaly or breakdown in understanding occurs, the individual strip is revisited with respect to the schema or "spirit of the whole," the one guiding idea that governs the text (Ormiston and Schrift 1990, p. 12). Through this dialectic process of reexamining strips and readjusting our schemas, we moved toward improved understanding of the whole text. **Replication**. A method used in case study research to support validity is replication. Through replication across multiple cases, the findings are shown to be generalizable beyond the immediate case…

Table 7 – formative criteria and summative criteria in the articles in the *MIS Quarterly* special issue on intensive research

	ARTICLE'S FORMATIVE CRITERIA	ARTICLE'S SUMMATIVE CRITERIA
Gopal, A. & Prasad, P. "Understanding GDSS in Symbolic Context: Shifting the Focus from Technology to Interaction," (24:3), 2000, p. 515.	Compatibility of research questions with symbolic interactionist assumptions and orientations Selection of appropriate field setting and methods Immersion Capture of multiple realities Familiarity with context(s) Thick description Maintaining narrative rather than scientific style Emphasizing the problematics of the research situation	
Nelson, K.M., Nadkarni, S., Narayanan, V.K. & Ghods, M. "Understanding Software Operations Support Expertise: A Revealed Causal Mapping Approach," (24:3), 2000, p. 484.	Research Focus: What are the objectives of research? Choice of Source: How do we capture relevant knowledge? Sampling Strategy: Does the choice of sample reflect research objectives? Construction of Maps: A. Categories: Are the categories conceptually relevant? B. Operationalization of constructs: How do we capture the concepts in a measurable manner? C. Operationalization of linkage: How do we capture the linkage among theoretical constructs? Unit of Analysis: Is the level of analysis consistent with the phenomena under investigation? Convergence: Is the knowledge structured of random? Validity of findings: do the findings make sense?	

Table 8 – Modus Tollens – four forms of academic inquiry

	Positivist Research	Interpretive Research	Action Research	Design Research
$p \supset q$	If a positivist theory about a phenomenon is true, then what the theory predicts about an instantiation of the phenomenon is true.	If a reader's interpretation of a text is a valid hermeneutic interpretation, then the reader's interpretation of a particular passage or set of passages in the text does not give rise to any contradiction, inconsistency, or other anomaly with regard to the reader's interpretation of any or all of the other passages in the text.	If a theory of action for solving a certain organizational problem is effective, then a particular action A prescribed by the theory of action for the organizational problem's instantiation P in organization O will solve the problem P.	If a design theory for solving a certain organizational problem is effective, then a particular artifact A prescribed by the design theory for the organizational problem's instantiation P in organization O will solve the problem P.
$\sim q$	For an instantiation of the phenomenon, what the theory predicts turns out not to be true.	For a particular passage or set of passages in a given text, the reader's interpretation gives rise to a contradiction, inconsistency, or other anomaly with regard to the reader's interpretation of another particular passage or set of other passages in the same text.	The particular action A prescribed by the given theory of action does not solve the organizational problem's instantiation P in organization O.	The particular artifact A prescribed by the design theory does not solve the organizational problem's instantiation P in organization O.
$\therefore \sim p$	Therefore the positivist theory is not true (thereby suggesting that an improved or new positivist theory needs to be developed and then also tested).	Therefore the reader's interpretation of the given text is not a valid hermeneutic interpretation (thereby suggesting that an improved or new interpretation needs to be developed and then also tested).	Therefore the statement that "the given theory of action for solving a certain organizational problem is effective" is not true or, simply, the given theory of action is not effective (thereby suggesting that an improved or new theory of action needs to be developed and tested).	Therefore the statement "the given design theory for solving a certain organizational problem is effective" is not true or, simply, the given design theory is not effective (thereby suggesting that an improved or new design theory needs to be developed and tested).

APPENDIX A

THE MAJOR PREMISE OF MODUS TOLLENS EXPRESSED IN DIFFERENT FORMS

major premise of modus tollens in everyday English	if *p* is true, then *q* is true
major premise of modus tollens in sentential or propositional logic	$p \supset q$
major premise of modus tollens in relational predicate logic with multiple quantifiers[20]	$(X)(Y)(Z)\Phi XYZ \supset \Phi xyz$
same as row above, using this essay's notation for independent and dependent variables in positivist research	$(Y)(X_1)(X_2)\ldots(X_{n-1})(X_n)\Phi YX_1X_2\ldots X_{n-1}X_n \supset \Phi yx_1x_2\ldots x_{n-1}x_n$
major premise of modus tollens in algebraic form with mathematical functions	If the theory's mathematically operationalized proposition "$Y = f(X_1, X_2, \ldots, X_n)$" is true, then the prediction (a statement describing an instantiation of the theory) "$y_{\text{predicted}}=f(x_1, x_2, \ldots, x_n)$" is true.
comments on row above	$f(x_1, x_2, \ldots, x_n)$ refers to the function $Y = f(X_1, X_2, \ldots, X_n)$ where the observed numerical value x_i for each independent variable X_i is plugged into the function. $f(x_1, x_2, \ldots, x_n)$ returns the calculated numerical value that the theory predicts for the dependent variable Y. $y_{\text{predicted}} \equiv f(x_1, x_2, \ldots, x_n)$. y_{new} is the actual, observed numerical value for the dependent variable Y. y_{new} is compared to $y_{\text{predicted}}$.

---

[20] A convention for expressing predicate logic designates variables with lower case letters and constants with upper case letters (Klenk 1983). Here, we are doing the opposite so as to follow a convention more familiar to some positivist researchers.

APPENDIX B

MODUS TOLLENS EXPRESSED IN DIFFERENT FORMS

	modus tollens in sentential or propositional logic	modus tollens in relational predicate logic with multiple quantifiers	modus tollens in relational predicate logic with multiple quantifiers, where the variables are interpreted as the traditional independent variables X_i and dependent variable Y that are familiar to positivist science	modus tollens in algebraic form with mathematical functions
major premise	$p \supset q$	$(X)(Y)(Z)\Phi XYZ \supset \Phi xyz$	$(Y)(X_1)(X_2)\ldots(X_{n-1})(X_n)\Phi YX_1X_2\ldots X_{n-1}X_n \supset \Phi yx_1x_2\ldots x_{n-1}x_nx_{\mathbf{y}}$	If "$Y=f(X_1,X_2, \ldots, X_n)$" is true, then the prediction "$y_{new} = f(x_1, x_2,\ldots,x_n)$" is true.
minor premise	$\sim q$	$\sim\Phi xyz$	$\sim\Phi yx_1x_2\ldots x_{n-1}x_nx_{\mathbf{y}}$	"$y_{new} = f(x_1, x_2, \ldots , x_n)$" is not true.
conclusion	$\therefore\sim p$	$\therefore\sim(X)(Y)(Z)\Phi XYZ$	$\therefore\sim(Y)(X_1)(X_2)\ldots(X_{n-1})(X_n)\Phi YX_1X_2\ldots X_{n-1}X_n$	Therefore "$Y = f(X_1, X_2, \ldots , X_n)$" is not true.

APPENDIX C

TESTING A THEORY WITH OUT-OF-SAMPLE DATA POINTS

To test a theory in its mathematically operationalized form, such as the relation "$BI=ß_0+ß_1A+ß_2U$," a researcher may proceed by combining the use of prediction intervals with the "leave one out" cross-validation technique, where the left-out data point plays the role of the out-of-sample data point and where each data point takes a turn at playing this role. If the theory is true, then a researcher would expect 95% of the predictions of BI to fall inside their respective 95% prediction intervals. However, how much would the percentage of successful predictions need to fall short of 95 before the researcher loses confidence in the theory and therefore rejects it as true?

Suppose that only 88 out of a total of 100 predictions fall inside their respective 95% prediction intervals. Does 88 fall sufficiently short of 95 for the researcher to make the decision, at the .01 or 1% level of statistical significance, to reject the theory?

To calculate the probability that, when the theory is true, the number of successful predictions reaches, at most, only 88, the researcher relies on the binomial distribution, where 1) the total number of trials, n, is 100, 2) the number of successful trials is s, and 3) the probability that any given trial is successful is 95%. Calculating the probability of reaching, at most, only 88 successful predictions involves, first, applying the binomial distribution 89 times, each time with a different value of s, where s ranges from 0 to 88, and then, second, adding up the 89 different probabilities. The fourth row of the first table below, "Testing 100 Predictions Using 95% Prediction Intervals," indicates that the 89 different probabilities add up to 0.004.

The number 0.004 means that, if the theory is true, there is only a 0.4% chance that the number of successful predictions will reach, at most, only 88. May 0.4% be considered small enough to doubt, and reject, the theory as true? If the researcher, prior to testing any predictions, chose .01 or 1.0% as the threshold level (i.e., the critical value), then the 0.4% chance would provide sufficient justification for making the decision to reject the theory as true. In other words, 1) if the “p-value” or simply “p” is 0.004, 2) if the threshold – the researcher’s desired level of statistical significance, α, which the researcher selected prior to the start of testing any predictions – is 0.010 or 1%, and 3) because $p=0.004$ is less than $\alpha=0.010$ (i.e., “$p<0.010$”), then the researcher has the justification needed to make the decision, at the .01 or 1.0% level of statistical significance, to reject the theory as true. In this case, *the theory does not survive empirical testing, and summative validity is not achieved.*

As already mentioned in the essay, “rejecting a theory” can call for replacing the *entire* theory with a completely new theory or a rival theory, but “rejecting a theory” can also mean something much less drastic: It can call for rejecting *just the current formulation* of the theory, which can involve making some adjustments in it (e.g., retaining most or all of the independent variables but changing the relationships between them, adding a new independent variable, introducing a moderating relationship, etc.).

For purposes other than those in this study, the second author collected data from a project financed by the Saudi Arabian government to assess factors that affect the acceptance and use of computers (as a technology) by knowledge workers in Saudi Arabia. The participating organizations represented various banking, merchandising, manufacturing, and petroleum industries. The survey solicited responses from professional knowledge workers in these organizations engaged in the use of desk top computers for the purpose of their work. Through this procedure, a total of 1,190 survey responses were collected. The survey collected data on three of the technology acceptance model’s constructs: perceived ease of use (EOU), perceived usefulness (U), and attitude (A). We now use the data to illustrate how to combine the use of prediction intervals with the “leave one out” cross-validation technique.

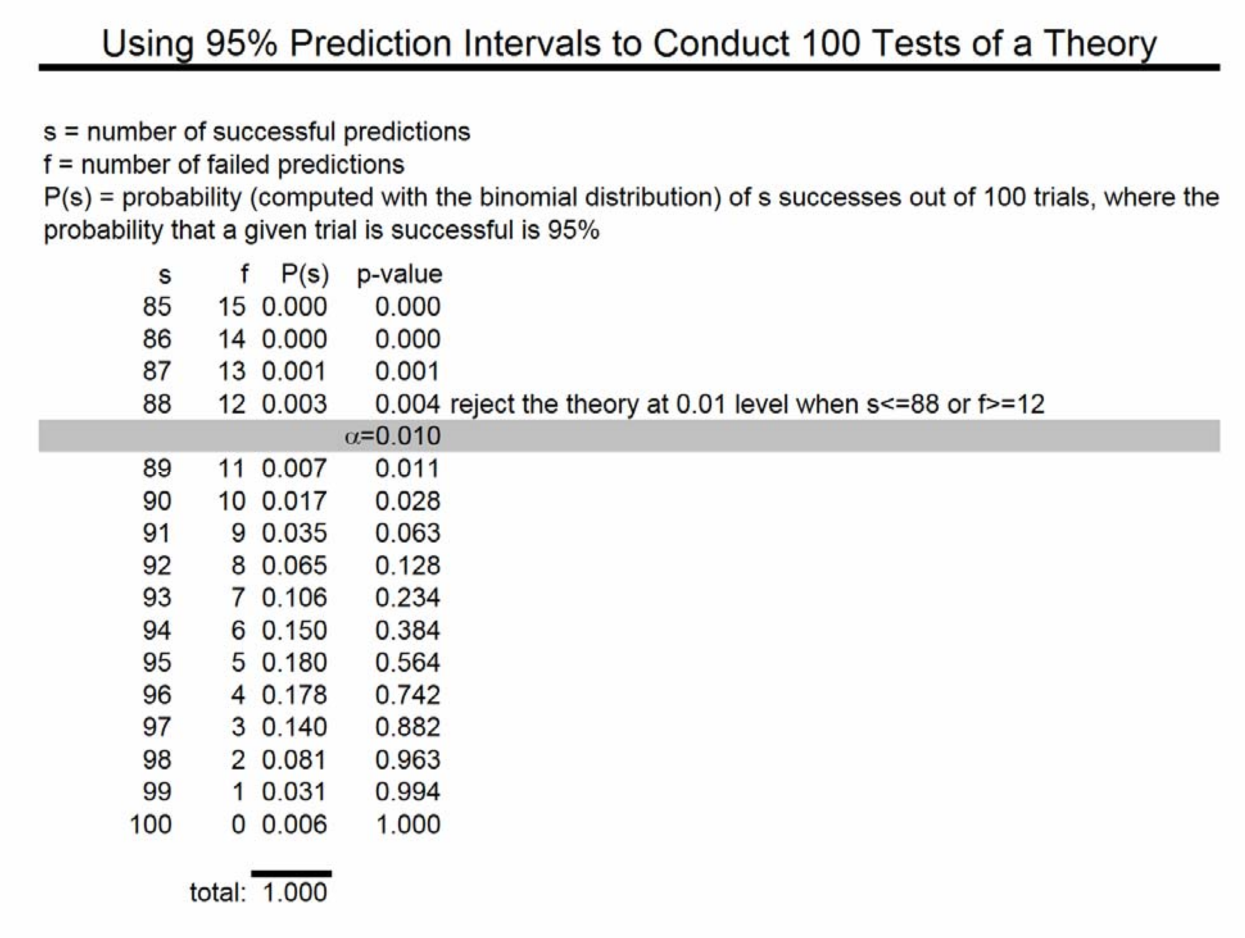
Suppose we use the technology acceptance model to play the role of the theory being tested. Suppose further that, for the sake of illustration, this theory is mathematically operationalized as just one relation, “$BI=ß_0+ß_1A+ß_2U$.” Combining the use of prediction intervals with the “leave one out” cross-validation technique, we let each and every one of the 1,190 samples points take a turn at playing the role of the out-of-sample data point. We thus conduct 1,190 experiments, each one of which tests the theory mathematically operationalized as “$BI=ß_0+ß_1A+ß_2U$.” It turns out that, in the 1,190 experiments, 1,121 of the predictions fall inside their 95% prediction intervals. For the 1,190 experiments, the expected number of predictions to fall inside their 95% prediction intervals is 1,131 (i.e., 95% of 1,190). Does reaching a level of only 1,121 successful predictions (which is 94% of all 1,190 predictions) instead of the expected number of 1,131 (which is 95% of all 1,190 predictions) provide sufficient justification for making a decision to reject the theory as true? Using the binomial distribution (where $n=1,190$ trials, s=number of successes, and the probability that a trial is a success is 0.95), we can calculate the probability of reaching, at most, 1,121 successful predictions when the relation “$BI=ß_0+ß_1A+ß_2U$” is true. This probability is $p=0.117$, which does not satisfy $p<.050$, much less $p<0.010$. Therefore, whether we use one or the other of these two customary levels of significance, our experimental result of reaching only 1,121 successful predictions or 94% does *not* provide sufficient justification for making a decision to reject the theory as true. Instead, the evidence is considered to be consistent with the theory. In this way, *we are successfully establishing the summative validity of the theory*. The evidence allows the researcher not to reject the theory, and to continue using it.

To check the effect of a smaller number of data points, we randomly select 258 data points from the total of 1,190. The result is 237 predictions which fall inside their 95% prediction intervals, for a 92% success rate. Using the binomial distribution (where n=258 trials, s=number of successes, and the probability that a trial is a success is 0.95), we can calculate the p-value associated with reaching, at most, 237 successful predictions out of a total of 258 predictions. The p-value turns out to be p=0.02, which is sufficiently small, when the desired level of statistical significance is 0.050 or the desired confidence level if 95%, to justify making the decision to reject the theory as true. *In this case, where the desired level of significance is 0.050 or 5%, the evidence indicates that the theory fails empirical testing and therefore lacks summative validity*.

Next, to see the effect of the number of experiments slightly larger than 258, we randomly select 301 data points from the original sample of 1,190. The result is 276 predictions which fall inside their 95% prediction intervals, for a 92% success rate. Using the binomial distribution (where n=301 trials, s=number of successes, and the probability that a trial is a success is 0.95), we calculate that the p-value which corresponds to 276 successful predictions is 0.001, which meets the more stringent desired level of statistical significance of α=0.010 or, equivalently, the desired confidence level of 99%. Because p=0.001, then “p<0.010” is true, *there is sufficient justification for making the decision to reject the theory as true and to conclude that the theory lacks summative validity*.

In the final repetition of this exercise, we randomly select 387 data points from the original sample of 1,190. 358 (93%) of them involve predictions that fall inside their 95% prediction intervals. Using the binomial distribution, we determine that the p-value for this result is 0.022, which is smaller than the desired significance level of α=0.050, but not when the desired significance level is $\alpha=0.010$. This means that we have sufficient evidence to justify making the decision, at the 0.050 significance level, but not the 0.010 significance level, to reject the theory as true and to conclude that the theory lacks summative validity.

An alternative to the laborious prediction-interval/leave-one-out method, which can involve testing hundreds or thousands of data points, is to test the predictions for only 10 data points that are randomly selected from all those in the original sample. When testing 10 predictions with 95% prediction intervals, a researcher may reasonably expect 10, 9, or even 8 of the predictions to fall inside their 95% prediction intervals if the theory being tested is true. As shown in the second table below, where a researcher is testing a total of just 10 (instead of 1,190) predictions, the experimental result in which only 6 successful predictions are reached (which is indicated in the row of the second table where “s<=6 or f>=4”) would justify making the decision, at the $\alpha=0.01$ significance level to reject the theory. And similarly, in the third table below, where a researcher is testing a total of just 5 (instead of 10) predictions, the experimental result in which only 2 successful predictions are reached (s<=2 or f>=3) would justify making the decision, at the 0.01 significance level to reject the theory as true.



Using 95% Prediction Intervals to Conduct 100 Tests of a Theory

s = number of successful predictions
f = number of failed predictions
P(s) = probability (computed with the binomial distribution) of s successes out of 100 trials, where the probability that a given trial is successful is 95%

s	f	P(s)	p-value	
85	15	0.000	0.000	
86	14	0.000	0.000	
87	13	0.001	0.001	
88	12	0.003	0.004	reject the theory at 0.01 level when s<=88 or f>=12
			α=0.010	
89	11	0.007	0.011	
90	10	0.017	0.028	
91	9	0.035	0.063	
92	8	0.065	0.128	
93	7	0.106	0.234	
94	6	0.150	0.384	
95	5	0.180	0.564	
96	4	0.178	0.742	
97	3	0.140	0.882	
98	2	0.081	0.963	
99	1	0.031	0.994	
100	0	0.006	1.000	
	total:	1.000		

Testing 10 Predictions Using 95% Prediction Intervals

s = number of successful predictions

f = number of failed predictions

P(s) = probability (computed with the binomial distribution) of s successes out of 10 trials, where the probability that a given trial is successful is 95%

s	f	P(s)	p-value	
0	10	0.0000	0.000	
1	9	0.0000	0.000	
2	8	0.0000	0.000	
3	7	0.0000	0.000	
4	6	0.0000	0.000	
5	5	0.0001	0.000	
6	4	0.0010	0.001	reject the theory at 0.01 level when s<=6 or f>=4
			α=0.010	
7	3	0.0105	0.012	
8	2	0.0746	0.086	
9	1	0.3151	0.401	
10	0	0.5987	1.000	
	total:	1.0000		

Testing 5 Predictions Using 95% Prediction Intervals

s = number of successful predictions

f = number of failed predictions

P(s) = probability (computed with the binomial distribution) of s successes out of 5 trials, where the probability that a given trial is successful is 95%

s	f	P(s)	p-value	
0	5	0.0000	0.000	
1	4	0.0000	0.000	
2	3	0.0011	0.001	reject the theory at 0.01 level when s<=2 or f>=3
			α=0.010	
3	2	0.0214	0.023	
4	1	0.2036	0.226	
5	0	0.7738	1.000	
	total:	1.0000		