Generalizing Generalizability in Information Systems Research

Allen S. Lee • Richard L. Baskerville
Virginia Commonwealth University, 1015 Floyd Avenue, Room 4146, Richmond, Virginia 23284-4000
Georgia State University, 33 Gilmer Street SE, Atlanta, Georgia 30303-3000
allenslee@alum.mit.edu • baskerville@acm.org

Generalizability is a major concern to those who do, and use, research. Statistical, sampling-based generalizability is well known, but methodologists have long been aware of conceptions of generalizability beyond the statistical. The purpose of this essay is to clarify the concept of generalizability by critically examining its nature, illustrating its use and misuse, and presenting a framework for classifying its different forms. The framework organizes the different forms into four types, which are defined by the distinction between empirical and theoretical kinds of statements. On the one hand, the framework affirms the bounds within which statistical, sampling-based generalizability is legitimate. On the other hand, the framework indicates ways in which researchers in information systems and other fields may properly lay claim to generalizability, and thereby broader relevance, even when their inquiry falls outside the bounds of sampling-based research.
(*Research Methodology; Positivist Research; Interpretive Research; Quantitative Research; Qualitative Research; Case Studies; Research Design; Generalizability*)

1. Introduction

Generalizability is a major concern to those who do, and use, research. Among other things, it refers to the validity of a theory in a setting different from the one where it was empirically tested and confirmed. A theory that lacks such generalizability also lacks usefulness. Because the field of information systems (IS) is not just a science but also a profession (and therefore has professional constituents such as IS executives, managers, and consultants), the generalizability of an IS theory to different settings is important not only for purposes of basic research, but also for purposes of managing and solving problems that corporations and other organizations experience in society. Statistical, sampling-based generalizability is a valid concept within its bounds, but its uncritical application as the norm for all generalizability can lead to an improper assessment of the generalizability of many research studies. The purpose of this essay is to clarify the concept of generalizability by critically examining its nature, illustrating its use and misuse, and offering a framework for classifying its different forms.

The *Oxford English Dictionary* (1989) defines generalize as "to form general notions by abstraction from particular instances," generalizable as "capable of being generalized," and generalizability as "the fact or quality of being generalizable." Conceptualized in this way, generalizability need not have a quantitative or statistical dimension. However, many IS researchers, both quantitative and qualitative, have restricted themselves to just one particular notion of generalizability—namely, a statistical, sampling-based notion. Furthermore, they have imposed this particular notion even outside the bounds of statistical, sampling-based research. It is as if statistical, sampling-based generalizability has been overgeneralized, as it were, to nonstatistical, nonsampling forms of research. Qualitative IS researchers have forgone claims to generalizability when, in fact, they have not yet broached conceptions of generalizability appropriate to their own research. Ultimately, this

1047-7047/03/1403/0221$05.00
1526-5536 electronic ISSN

notion of generalizability eliminates access to the insights that many information systems researchers offer in their research findings.

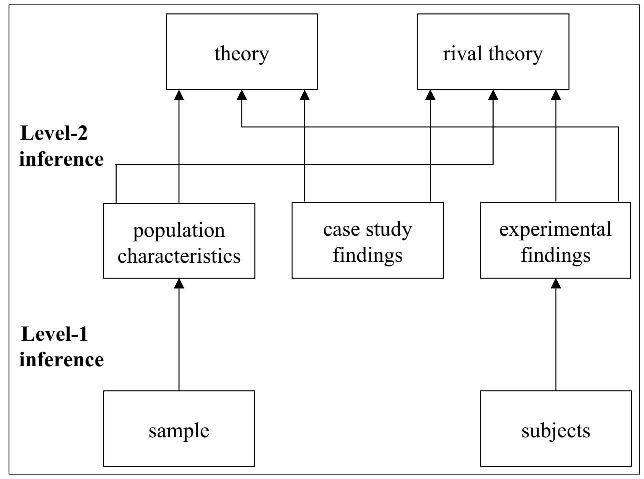
Our purpose in this methodological essay is to explicate different types of generalizability. One of our intended contributions is to make the different types of generalizability distinct and visible to researchers, thereby encouraging them to provide explicit treatment of the type(s) of generalizability (not just statistical) that they should identify as appropriate to their particular studies. Another intended contribution is to provide guidance to reviewers, editors, and authors in their efforts to apply generalizability arguments that are appropriate to the authors' (not just the reviewers' or editors') research philosophy and purpose. The framework that we present for distinguishing four types of generalizability introduces concepts that do not displace, but supplement, the well-known and widely accepted concepts associated with statistical, sampling-based research. Finally, a clarification of the different meanings and forms of generalizability can help to clarify how one might apply IS research findings to actual organizational settings apart from those where the research took place, and thereby promote the relevance of the research to professional practice.

In the second and next section of this essay, we describe the misapplication (or perhaps, the overapplication) of the concept of statistical, sampling-based generalizability in IS research. After that, we offer a critique of statistical, sampling-based generalizability. In the essay's fourth section, we examine alternatives to the statistical perspective on generalizability. Following that, we present the framework of four types of generalizability, and we also provide a discussion and illustration of it.

2. Misapplication of the Concept of Statistical Generalizability in IS Research

Methodologists have long been aware of conceptions of generalizability beyond the statistical. One such methodologist is Yin, whose case study methodology (1984, 1994) has received recognition from many IS researchers. Yin (Figure 1) describes the process

Figure 1 Yin's Conception of Generalization



Note. This illustration is a quotation of Figure 2.2 in Yin 1994, p. 31.

of generalizing from a sample to population characteristics or from experimental subjects to experimental findings as a form of what he calls Level-1 inference. He distinguishes this level from the subsequent and separate research activity of generalizing from the population characteristics to theory or from the experimental findings to theory; this activity is a form of generalization that he calls Level-2 inference. (He also refers to Level-2 inference as analytical generalization.) Yin explains that generalizing from case study findings to theory is a form of Level-2 inference. In this view, statistical generalizability is a measure of the quality of only one form of Level-1 inference. As such, statistical generalizability is inappropriate as a measure of the quality of case studies because they involve not only a different form of inference (inference from case study findings to theory rather than from a sample to population characteristics), but also inference at a different level (Level 2 rather than Level 1). However, despite the availability of Yin's and other conceptions of generalizability (which we will examine later), the statistical, sampling-based conception of generalizability remains widely and inappropriately used in nonstatistical, nonsampling research associated with top IS journals and conferences. As the research examples in Table 1 make evident, the authors of the research examples have transferred, from statistical research to

Table 1 Published Research that Applies the Statistical, Sampling-Based Conception of Generalizability to Nonstatistical, Nonsampling Research: Examples from Case Research

"There are, of course, many methodological limitations of this study. As a *small sample-size, single* case study, *generalizability cannot be assessed* . . . ." Majchrzak, A., R. Rice, A. Malhotra, N. King, B. Sulin. 2000. Technology adaptation: The case of a computer-supported interorganizational virtual team. *MIS Quart.* **24**(4) 569–600. (Winner of the award for *MIS Quart.* Best Paper of the Year 2000.)

"To what extent can the findings from this study generalize to other organizations and their practice of SM? . . .This study has some limitations. . . .While the case study design provided us with rich data from multiple sources (i.e., interviews, questionnaires, electronic logs, observations), the qualitative nature of the study does not lend itself to *rigorous statistical* analysis and causal inference. Finally, the *sample* was relatively *small*, involving *five* organizations and *eight* cases." Tyran, C., A. Dennis, D. Vogel. 1992. The application of electronic meeting technology to support. *MIS Quart.* **16**(3) 313–334.

"First and foremost, it should be reaffirmed that the *single case* research strategy employed here *only allows* generalizability to a research model, which in turn needs to be tested under a multiple case study design or by other field methods." Brown, C. 1997. Examining the emergence of hybrid IS governance solutions: Evidence from a single case site. *Inform. Systems Res.* **8**(1) 69–94.

"We *must not generalize* from a *single* case study in *one* organization." Vandenbosch, B., M. Ginzberg. 1996. Lotus notes and collaboration: Plus ça change . . . ." *J. Management Inform. Systems* **13**(3) 65–81.

"While a *single-site* case study *limits the ability to generalize*, there is support for the result that a well-executed information system based on ABC principles can improve management decision making and organizational performance." Stuchfield, N., B. Weber. 1992. Modeling the profitability of customer relationships: Development and impact of Barclays de Zoete Wedd's BEATRICE. *J. Management Inform. Systems* **9**(2) 53–76.

"The findings of this paper may also be *limited* because the paper is based on *only one case study*. Even though this case study was conducted longitudinally and six major IS decisions made over a 17-year period were examined, *it is very difficult to generalize this study's results to other organizations*." Newman, M., R. Sabherwal. 1996. Determinants of commitment to information systems development: A longitudinal investigation. *MIS Quart.* **20**(1) 23–54.

"From the point of view of theory development, while case studies provide useful anecdotal information, the *generalizability from one specific instance* to another is often *limited*." Albers, M., R. Agarwal, M. Tanniru. 1994. The practice of business process reengineering: Radical planning and incremental Implementation in an IS organization. *SIGCPR '94. Proc. 1994 Comput. Personnel Res. Conf. Reinventing IS: Managing Inform. Tech. Changing Organ.*, ACM Press, Arlington, VA, 87–96.

"From the evidence of the *two cases*, it was *not possible to identify any generalisable strategies* for overcoming constraints but the particular solutions developed appeared to reflect the developers' local conditions and their knowledge, intuition, and experience. This would suggest that rather than giving a set of generalized guidelines for improving user involvement (as is common in the literature), the emphasis might be better placed on supporting developers' ingenuity and improvisation and on developing their social skills to enable them to overcome the constraints on involvement." Nandhakumar, J., M. Jones. 1997. Designing in the dark: The changing user-developer relationship in information systems development. K. Kumar, J. DeGross, eds. *Proc. 18th Internat. Conf. Inform. Systems*, 75–87. (Winner of the 1997 ICIS Best Paper Award.)

"Because they are drawn from a study of *two* organizations, these results *should not be generalized* to other contexts. Each context is different, so we should expect different contextual elements to interact with technical initiatives to produce different consequences. The findings should not even be extended to other settings where GIS, or even Arc/Info, is implemented. What is true for GIS in the two local county governments studied may be untrue for GIS in other governmental units or in private enterprises." Robey, D., S. Sahay. 1996. Transforming work through information technology: A comparative case study of geographic information systems in county government. *Inform. Systems Res.* **7**(1) 93–110.

"In particular, in-depth analysis of extensive data *from only one organization reduces generalizability*, but increases correspondence to reality." Hidding, G. 1998. Adoption of IS development methods across cultural boundaries. R. Hirschheim, M. Newman, J. DeGross, eds. *Proc. 19th Internat. Conf. Inform. Systems*, 308–312.

"The study has a number limitations that need to be considered in making any conclusions. First, *the single case site limits the generalizability of results*. The purpose of the study was not to provide generalizability of empirical results to other firms, rather the purpose was to 'expand and generalize theories' " (Yin 1984). Jarvenpaa, S. L., D. Leidner. 1997. An information company in Mexico: Extending the resource-based view of the firm. K. Kumar, J. DeGross, eds. *Proc. 18th Internat. Conf. Inform. Systems*, 75–87.

"Conducting additional case studies will provide instances of the various learning/outcome combinations, and we encourage such research. On the other hand, *case studies alone will not result in validity or generalizability*. Toward that end, a more fruitful approach might be to compare the development processes for similar systems in different organizations, or two or more systems being developed in a single organization." Stein, E., B. Vandenbosch. 1996. Organizational learning during advanced system development: Opportunities and obstacles. *J. Management Inform. Systems* **13**(2) 115–136.

qualitative research, both the notion of sampling and the associated notion that a small sample size (e.g., only one organization) limits generalizability.

One possible explanation for these remarks is that the reviewers, editors, and program chairs for these authors have placed pressure on them to conform to the statistical, sampling-based conception of generalizability as the norm. Another possible explanation is that generalizations are sometimes mistakenly expected to be proven statements, rather than taken as well-founded but as-yet untested hypotheses. Whatever the explanation, the result is that authors themselves (and their reviewers) have misapplied or overapplied the concept of statistical generalizability rather than taken advantage of other conceptions of generalizability appropriate to the authors' inquiry. By extending generalizability, authors may better draw specific attention to the relevance of their findings. Before we review additional conceptions of generalizability, we offer a critique of statistical generalizability.

3. A Critique of Statistical, Sampling-Based Generalizability

Statistical, sampling-based generalizability is a valid concept, but its uncritical application as the norm for all generalizability can lead to an improper assessment of the generalizability of a research study. To critique statistical generalizability, we will first examine inductive reasoning, of which statistical generalizing is a form. The examination will focus on Hume's truism, which calls attention to an irremediable problem in induction. Second, we will examine what Hume's truism means for statistical generalizability.

Hume's Contribution to the Understanding of Induction

Induction refers to a process of reasoning and can be a synonym for generalizing. It refers to a reasoning process that begins with statements of particulars and ends in a general statement. Reasoning from data points in a sample to an estimate of a population characteristic is an instance of induction. Campbell and Stanley call attention to "some painful problems in the science of induction" (1963, p. 17, original emphasis retained):

> The problems are painful because of a recurrent reluctance to accept Hume's truism that *induction or generalization is never fully justified logically*. Whereas the problems of *internal* validity are solvable within the limits of the logic of probability of statistics, the problems of external validity are not logically solvable in any neat, conclusive way. Generalization always turns out to involve extrapolation into a realm not represented in one's sample. Such extrapolation is made by *assuming* one knows the relevant laws. Thus, if one has an internally valid Design 4,[1] one has demonstrated the effect only for those specific conditions which the experimental and control group have in common, i.e., only for pretested groups of a specific age, intelligence, socioeconomic status, geographic region, historical moment, orientation of the stars, orientation of the magnetic field, barometric pressure, gamma radiation, etc.
>
> *Logically*, we cannot generalize beyond these limits; i.e., we cannot generalize at all. But we do attempt generalization by guessing at laws and checking out some of these generalizations in other equally specific but different conditions. In the course of the history of a science we learn about the "justification" of generalizing by the cumulation of our experience in generalizing, but this is not a logical generalization deducible from the details of the original experiment. Faced by this, we do, in generalizing, make guesses as to yet unproven laws, including some not yet explored . . . .

Hume, an 18th century Scottish philosopher, "is almost universally credited with discovering the problem of induction" Rosenberg 1993, p. 75). Wood (2000) offers a detailed explanation of Hume's problem of induction. The problem of induction is about how to establish induction itself as a valid method for empirical inquiry.

Induction can be expressed in the form of Argument 1.1 in Figure 2. The status of induction as a valid method of empirical inquiry is open to question because the second statement does not logically follow from the first. Wood refers to this as Problem 1. Wood continues: "To make Argument [1.1] valid, we

[1] Campbell and Stanley describe "Design 4" as follows (1963, p. 13):

$$\begin{array}{llll} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \end{array}$$

Figure 2 First Attempt to Justify Induction

Argument 1.1
- In past experience, all *F*s have been *G*s.
- Therefore, the next *F* will be a *G* or all future *F*s will be *G*s.

Argument 1.2
- If in past experience, all *F*s have been *G*s, then the next *F* will be a *G* or all future *F*s will be *G*s.
- In past experience, all *F*s have been *G*s.
- Therefore, the next *F* will be a *G* or all future *F*s will be *G*s.

Note. Based on Wood (2000).

Figure 3 Second Attempt to Justify Induction

Argument 2.1
- In past experience, all tests have confirmed Theory 1.
- Therefore, the next test will confirm Theory 1 or all future tests will confirm Theory 1.

Argument 2.2
- If in past experience all tests have confirmed Theory 1, then the next test will confirm Theory 1 or all future tests will confirm Theory 1.
- In past experience, all tests have confirmed Theory 1.
- Therefore, the next test will confirm Theory 1 or all future tests will confirm Theory 1.

Note. Based on Wood (2000).

need an additional premise, such as [the] Uniformity of Nature assumption or: 'The future will be like the past'," where the result is Argument 1.2.

Argument 1.2 employs a form of the uniformity of nature assumption as the first statement in an argument that takes the form of a syllogism, which consists of a major premise, minor premise, and conclusion. The major premise is the first statement in the syllogism. The second statement, "In past experience, all *F*s have been *G*s," plays the role of the minor premise. Applying the major premise to the minor premise leads deductively to the conclusion, "Therefore, the next *F* will be a *G* or all future *F*s will be *G*s." Note that the conclusion in Argument 1.2 is the same as the second statement in Argument 1.1. Therefore, if Argument 1.2 were valid, it would provide a proper way of establishing the validity of induction.

Whereas Argument 1.2 performs its deductive reasoning correctly, the conclusion in any syllogism can be valid only if its major premise is valid. Wood refers to the following as Problem 2: In Argument 1.2, how would we know that the major premise—the uniformity of nature proposition—itself is valid? We would therefore need to take a step back in order to establish the validity of the uniformity of nature premise.

Wood explains that there are two ways by which we could attempt to establish the validity of the uniformity of nature proposition, which is denoted as Theory 1 in Figure 3. One way is by recourse to Argument 2.1, but its mode of reasoning is induction exactly as Argument 1.1's mode of reasoning was induction; therefore, the same Problem 1 that arose for Argument 1.1 would also arise for Argument 2.1. To remedy this instance of Problem 1, we would again need an additional premise, where the result is Argument 2.2.

As it turns out, Argument 2.2 employs the uniformity of nature proposition as its major premise, just as Argument 1.2 did. Because Argument 2.2 takes the form of a syllogism, its conclusion can be valid only if its major premise is valid. The result is that Problem 2 would recur: How would we know that the major premise in Argument 2.2 is valid? We would need to take a step back in order to establish the validity of the major premise in Argument 2.2, just as we previously took a step back in order to establish the validity of the major premise in Argument 1.2. The result is that we would find ourselves in an infinite regress taking the form of Figures 4, 5, and so on, where the stream of reasoning would have no conclusion. Rosenberg offers a succinct description of Hume's truism (1993, p. 75):

> Hume recognized that inductive conclusions could only be derived deductively from premises (such as the uniformity of nature) that themselves required inductive warrant, or from arguments that were inductive in the first place. The deductive arguments [e.g., Arguments 1.2 and 2.2] are no more convincing than their most controversial premises and so generate a regress, while the inductive ones [e.g., Arguments 1.1 and 2.1] beg the question. Accordingly, claims that transcend available data, in particular predictions and general laws, remain unwarranted.

The enormous significance of Hume's truism leads Campbell and Stanley (1963) to take the positions that "*induction or generalization is never fully justified logically*" and that "we cannot generalize at all" (emphasis in the original, cited above). However, they recognize that a "recurrent reluctance" among scientific researchers to accept Hume's truism remains in force. Table 1 (above) demonstrates that, among IS researchers, this reluctance remains in force today. Specifically, they concede that a small sample size or small number of cases will offer only limited or no support to the generalization. The gathering of additional evidence can indeed be beneficial to the reliability of a scientific study. However, as we will describe in the next section, this is distinct from any increase in the generalizability of a sample to its population. The remainder of this section will apply Hume's truism to statistical generalizability in order to show what statistical generalizability does, and does not, properly involve.

The Ramifications of Hume's Truism for Statistical Generalizability

Hume's problem of induction can help to clarify the fundamental limitations of statistical, sampling-based generalization and the accompanying issue of how sample size is, and is not, related to generalizability. An increase in the size of a random sample means that sample-based estimates will increase in reliability. However, this does not mean that those estimates will become better generalizable to any population characteristics. For example, repeated sampling from the same population will result in sample estimates (e.g., the average for each sample) whose numerical values are closer to each other or more convergent than for smaller-sized samples. Where a sample point refers to a datum in the sample, and where an increase in the size of the sample means that it contains more such data, one can say that an increase in sample size will also lead to greater *generalizability of the sample points to a sample estimate* because of the greater convergence expected from the larger sample size. Note, however, that this is distinct from the *generalizability of the sample estimate to the corresponding population characteristic*. The latter generalizability, according to Hume's truism, cannot be improved (or, for that matter, even established) by increasing the sample size. The following discussion will elucidate these points.

Consider the null hypothesis, H_0, that the average of the perceived usefulness, U, which a population of managers associates with a particular technology (where U is measured on a scale from 1 to 5), is 3.[2] In other words, a researcher does not know the numerical value of the average of U for the population, but hypothesizes it to be 3. The researcher then tests the hypothesis, H_0, by taking a random sample of managers from the population, where the researcher uses the average of U for the sample as an estimate of the average of U for the population. Taking the sample randomly from the population is important for helping to ensure that the sample turns out to be representative of the population. The basic idea is that an average of U close to 3 for a random sample would be evidence consistent with the hypothesis that the average of U for the population is 3, while a sample average quite different from 3 would be evidence calling for the researcher to reject this hypothesis.

If the legitimacy of inductive reasoning is unquestioned, then the researcher could reason that the sample average is generalizable to the population average and that an increase in the size of a sample yielding an average of 3 would provide even stronger evidence for the generalization that the average of the population is 3. However, if the legitimacy of inductive reasoning is questioned and Hume's argument is applied, there would be no sound basis for making any statement about the value of the population average. Using Wood's terminology, we may express this as follows: Just because all Fs in past samples have averaged out to 3 does not mean that all or any Fs remaining to be observed in the population will average out to 3 or, for that matter, even anything close to 3. Therefore, an increase in sample size will not increase the generalizability of the average of U in the sample to the average of U in the population. In fact, Hume's elucidation of the problem of induction would prohibit a researcher from claiming that a sample of any size provides support for the generalizability of any sample estimate to the corresponding population characteristic.

An increase in sample size is beneficial, but the benefits take the form of improved *reliability of the sampling procedure*, rather than improved *generalizability of a sample to its population*. Reliability refers to the extent to which the same or a different researcher can, at least in principle, reapply the same procedure when making another observation or measurement

[2] The null and alternative hypotheses are H_0, $\mu_U = 3$ and H_1, $\mu_U \neq 3$.

Figure 4 Statistical Significance as a Conditional Probability

statistical significance $= \alpha$

$= P(\text{decision error} \mid H_0)$

level of confidence $= 1 - \alpha$

Increasing the sample size will improve the numerical value of what is synonymously known as α and $P(\text{decision error} \mid H_0)$.

This is an assumption: "on the condition that," "assuming that," "given that," or "in the event that" H_0 is true. In our example, H_0 is the hypothesis that the average of the population is 3.

(e.g., observing the average of another random sample of the same size taken from the same population) and expect it to lead to the same result as before (where the result is either to reject or not to reject the null hypothesis). An increase in the sample size will lead to an increase in reliability, as we explain in the following.

A 95% confidence level and a level of statistical significance at 0.05 are equivalent ways by which statisticians denote the following idea about reliability: In the event that the null hypothesis is true (e.g., the hypothesis that the average of U for the population is 3), the researcher may expect, for every 100 random samples of the same size that could be taken from the population, 95 of them to be sufficiently representative of the population (e.g., samples whose values for the average of U are close to 3) so as to lead to the correct decision of not rejecting the null hypothesis. For the 100 random samples in this example, an increase in the size of each sample would have the benefit of increasing—from 95 to, say, 98—the expected number of samples that lead to the correct decision of not rejecting the null hypothesis. In other words, an increase in the sample size will increase the reliability of the sampling procedure. An equivalent way of stating this is that, for a sample (one of the 100 in our example) that leads to the correct decision of not rejecting the null hypothesis, an increase in sample size would result in a larger number of other samples to which this sample could itself be generalized. Therefore, a larger sample size does increase generalizability, but it is the generalizability of a sample to other samples, not to the population.

We note that any increase in the generalizability of one sample to other samples would make no difference to the status of the null hypothesis as true or false. Its being true or false would still remain the same (and, to the researcher, unknown). Indeed, the confidence level, p-value, or statistical significance can be computed only for the portion or subset of the universe where the null hypothesis is assumed to be true, as Figure 4 indicates in the form of the condition in a conditional probability. An increase in the sample size would make a difference to the level of statistical significance or $P(\text{decision error} \mid H_0)$, but would make no difference to the certainty with which H_0 is known to be true, $P(H_0)$.

This line of reasoning holds that an increase in the size of a random sample will, if repeated samples are taken, result in sample estimates (e.g., the average for each sample) whose numerical values are more convergent than in the situation where the sample size is not increased. Thus, the increase in sample size improves the generalizability of sample points to a sample estimate. However, this is distinct from the sample estimate's generalizability to the population characteristic of which it is an estimate. An increase in sample size makes no difference to the latter generalizability.

Summary of the Critique of Statistical Generalizability

The main ramification of Hume's truism for statistical, sampling-based generalizability is that it prohibits the conclusion that an increase in the size of a sample leads to an increase in the generalizability of any sample estimate to its corresponding population characteristic. However, there is no prohibition of the conclusion that an increase in sample size leads to an increase in the generalizability of one sample to other samples that the same sampling procedure would produce. As a form of reliability, this conception of statistical, sampling-based generalizability is consistent with the traditional meaning of statistical significance (α) or the confidence level ($1-\alpha$): An increase in the confidence level exactly denotes an increase in reliability as we have described it. In fact, a caution that teachers traditionally express to students taking their first statistics course is an expression of exactly this meaning; it is the caution that, for instance, a 5% level of statistical significance or a 95% confidence level does *not* mean that the null hypothesis has a 95% probability of being true, but instead means that, *in the event that the null hypothesis is true*, the same sampling procedure would lead us to make the correct decision of not rejecting the null hypothesis 95% of the time, and hence the incorrect decision of rejecting it 5% of the time. The latter decision is also what statisticians call a "Type I" error. Hence our application of Hume's truism does not diminish, but affirms, proper statistical reasoning.

The conclusion to our critique is that we affirm the legitimacy of statistical generalizability, where we emphasize that it refers to the generalizability of one random sample to other random samples that would result from applying the same sampling procedure to the same population or the generalizability of sample points to a sample estimate. Given the purpose of forming general notions by abstraction from particular instances, statistical generalizability hardly constitutes a general model for all forms of generalizability. The singular nature of statistical generalizability therefore motivates an investigation into additional forms that generalizability can take.

4. Alternatives to the Statistical Perspective on Generalizability

Looking beyond statistical generalizability, one can observe that different researchers and philosophers have used the term generalizability to mean different things. In this section of the essay, we will examine different conceptions of generalizability in two different philosophical traditions.

Positivism and interpretivism are two major philosophical traditions that have received widespread attention from scientific researchers in the IS discipline (Boland 1985; Lee 1991, 1999; Lee et al. 1997; Mumford et al. 1985; Nissen et al. 1991; Orlikowski and Baroudi 1991; Trauth 2001; Walsham 1993). Statistical, sampling-based generalizability falls under the heading of the philosophical tradition of positivism. An examination of the two philosophical traditions will allow us to identify forms of generalizability in addition to the statistical.

For the most part, positivism and interpretivism have long been regarded as being in opposition and, hence, as supporting different positions on generalizability. Schutz,[3] a philosopher and phenomenologist, observed in 1962 that a controversy "for more than half a century has split not only logicians and methodologists but also social scientists into two schools of thought," where the schools of thought are today called positivism and interpretivism. He offers a succinct formulation of the controversy (1962–1966, p. 48):

> ...One [school of thought, positivism] holds that the methods of the natural sciences which have brought about such magnificent results are the only scientific ones and that they alone, therefore, have to be applied in their entirety to the study of human affairs....
>
> The other school of thought [interpretivism] feels that there is a basic difference in the structure of the social world and the world of nature. This feeling led to the other extreme, namely, the conclusion that the methods of the social sciences are *toto coelo* different from those of the natural sciences.... It has been maintained that the social sciences are idiographic, characterized by individualizing conceptualization and seeking singular assertory propositions,

[3] Schutz's perspective is significant because of the influence he has had on the practice of social scientific research: For example, his students Berger and Luckmann presented Schutz's phenomenological ideas to a social science audience in their book *The Social Construction of Reality* (1966), his student Garfinkel established ethnomethodology (1967), and the critical theorist Habermas relied on Schutz in conceptualizing his (Habermas') own ideas on the logic of the social sciences (Habermas 1988).

> whereas the natural sciences are nomothetic, characterized by generalizing conceptualization and seeking general apodictic propositions....

At the extreme, positivism would hold that any field of study, in order to qualify as scientific, must be nomothetic and therefore work towards the ideal of discovering universal or general laws; at the same time, interpretivism would hold that the goal of universal laws is inappropriate in the study of human affairs because individuals, groups, and other social units are all unique, and therefore demand idiographic theorizing instead.[4]

The controversy, already a half-century old when Schutz observed it in 1962, has not been resolved. Burrell and Morgan (1979) note the contrasting ontologies and epistemologies between what they call objectivism and subjectivism (a dualism that parallels positivism and interpretivism), and Hirschheim and Klein (1989) apply Burrell and Morgan's distinctions to the IS field. Orlikowski and Baroudi (1991) also recognize the contrast between positivism and interpretivism while noting the former's predominance in IS research. Lee (1991) recognizes the two traditions as well. He discusses them in his attempt to convey, to organizational and IS researchers, Schutz's solution of how to forge a rapprochement between positive and interpretive research. This is a rapprochement that Walsham (1995a, p. 382) believes "would be strongly opposed by some supporters of interpretivism on the grounds that [Lee's views] confuse and conflate contradictory epistemological positions." Hence, the controversy remains alive. To the extent that positivist researchers and interpretive researchers subscribe to the belief that their respective traditions use constructs and methods that are *toto coelo* different from each other's, the two groups will not see generalizability in the same light. In order to describe the details of the contrast between positivist and interpretive conceptualizations of generalizability, it is appropriate for us first to examine some selected fundamentals of the two philosophical traditions.

Capturing the spirit of positivism is a description of it as the natural-science model of social science. This description includes the following elements: (1) Inquiry in the natural sciences, especially physics and astronomy, provides the model that inquiry in the social sciences needs to follow; (2) there is an objective reality or real world that exists independently of scientific researchers; (3) there is an emphasis on the principle that, in science, "all knowledge regarding matters of fact is based on the 'positive' data of experience" rather than on mere opinion, speculation, or other unverifiable beliefs (Feigl 2002); and (4) the process of developing a scientific theory employs hypothetico-deductive logic, which allows a scientist to craft the theory's propositions so as to be not only logical (i.e., consistent with one another, which is best achieved by expressing the theory mathematically), but also empirical (i.e., faithfully portraying the real world, which is best achieved by testing the theory experimentally). To test the empirical validity of a theory, hypothetico-deductive logic employs the deductive logic of the syllogism, in contrast to inductive logic. Additional data collected in observations of additional portions of the real world would, in the hypothetico-deductive development of a theory, entitle it to a claim of greater generalizability (specifically, generalizability to the additional portions of the real world where the theory has actually undergone successful empirical testing).

An emphasis on generalizability is a key feature of the philosophical tradition of positivism. Positivism's "sole aim is to discover invariable universal laws governing phenomena" and it "seeks to determine the universal laws governing every observed phenomenon" (Kolakowski 1968, pp. 58–59); here, "universal" is a synonym for "general" or "generalizable." Also, "the natural sciences are nomothetic, characterized by generalizing conceptualization" (Schutz, quoted above). Combined with the positivist requirement that the social sciences must model themselves on the natural sciences, the positivist account of science as nomothetic would require inquiry in the social

[4] Luthans and Davis (1982, p. 380) quote the psychologist Gordon Allport as having borrowed this distinction in 1937 from neo-Kantian philosophy: "The nomothetic [sciences]...seek only general laws and employ only those procedures admitted by the exact sciences. Psychology in the main has been striving to make of itself a complete nomothetic discipline. The idiographic sciences...endeavor to understand some *particular* event in nature or in society. A psychology of individuality would be essentially idiographic."

sciences to strive for, as one of its goals, generalizability in the form of universal laws of human affairs.

In contrast, the philosophical tradition of interpretivism places no particular emphasis on generalizability or the striving for universal laws. In interpretivism, a theory's pertaining only to the setting where it was developed would not detract from its validity or scientific status. At the same time, interpretivism would not prohibit the researcher from extending his or her theory to additional settings.

A key feature of interpretivism that differentiates it from positivism, and hence also differentiates its approach to generalizability from positivism's approach, is that interpretivism acknowledges the existence of a phenomenon that is not present in the subject matter studied by the natural sciences. People, who are integral to the subject matter that a social scientist observes, develop and use their own subjective understandings of themselves, their setting, and their history. Therefore, already present in the subject matter of the social sciences are the meanings that people create and that they attach to the world around them. In this sense, subjective meaning is objective reality: The meanings that human subjects create, communicate, and hold are part and parcel of the real world that a social scientist receives as the subject matter under investigation. The presence of humanly created, and therefore sometimes contradictory, meanings and socially constructed realities in the subject matter of the social sciences has no counterpart in the subject matter of the natural sciences: "The world of nature, as explored by the natural scientist, does not 'mean' anything to molecules, atoms, and electrons" (Schutz 1962–1966, p. 59).

This feature of the subject matter of the social sciences leads to the distinction between "first-level constructs" and "second-level constructs" (Schutz 1962–1966, Van Maanen 1983).[5] For example, in ethnography, Van Maanen explains that first-level constructs "are the 'facts' of an ethnographic investigation" (p. 39) where the facts include "the situationally, historically, and biographically mediated interpretations used by members of the organization to account for a given descriptive property" (p. 40). In other words, the first-level constructs refer to the understandings held by the observed people themselves. In contrast, the second-level constructs refer to the understanding held by the observing researcher. Second-level constructs "are those notions used by the fieldworker to explain the patterning of the first-order data" and can include "statements about relationships between certain properties observed to covary in the setting" (p. 40). Van Maanen describes second-level constructs as "interpretations of interpretations" (p. 40) because they are the researcher's constructs of the first-level constructs. He is also explicit about labeling the researcher's constructs as theory: "The second-order concepts are the 'theories' an analyst uses to organize and explain these [first-level] facts" (p. 39); "theories developed by ethnographers in the field have an alterable and fluid character to them" (p. 51); and "theories are tested, retested, and tested again in the field" (p. 51).[6]

Schutz draws attention to the condition that the second-level constructs—i.e., theory—need to satisfy the requirements of science (pp. 62–63): "The constructs formed by the social scientist are constructs

[5] Van Maanen uses the equivalent terms "first-order concepts" and "second-order concepts."

[6] Empirical statements can be viewed as objective or subjective. The interpretive scholars Schutz (1962–1966), Geertz (1973), and Van Maanen (1983) all agree that interpretive research involves theory—second-level constructs—and not just the observing researcher's description of the observed people's first-level constructs. The first-level constructs (the facts) that the researcher interprets and records in order to provide a thick description of the lives of the observed people are, in a sense, just a passing instantiation of the people's culture and social structure. In contrast, the second-level constructs (theoretical formulations fashioned by the observing scientist) are necessarily more general than the descriptive portrayal of an instantiation; Geertz asserts categorically (p. 25), "if they are not general, they are not theoretical." Hence, generalizability is an essential feature of interpretive research that endeavors to provide theory and not just description. Insofar as the interpretive researcher begins with statements of particulars (the facts) and ends with a general statement (the theory), this reasoning process is a form of generalization where this is, in fact, precisely what Yin calls analytic generalization (1994, quoted above). However, unlike positivist generalizability whose ideal is universal laws, interpretive generalizability seeks to formulate theory so that it not only explains what the fieldworker has already observed, but also might help the same or different researcher to anticipate, or at least not be surprised by, additional observations subsequently made in the same setting (Sanday 1983, p. 22).

of the constructs formed in common-sense thinking by the actors on the social scene ... The scientific constructs formed on the second level ... are objective ideal typical constructs and, as such, of a different kind from those developed on the first level of common-sense thinking which they have to supersede ...." Along the same lines, Van Maanen adds (p. 41): "Second-order conceptions are relevant primarily to the culture of the researcher, not the researched," where the culture of the researcher includes what Schutz (p. 59) calls "the procedural rules of his [the researcher's] science."

The distinction between first-level constructs and second-level constructs calls for the social sciences to adopt methods different from (not necessarily instead of, but in addition to) those used in the natural sciences. For a researcher who seeks to examine, for instance, the meaning that the immediate setting has for the natives in a village or the managers in a corporation, the researcher employs methods suited to the observation of first-level constructs, which are methods for which there is no need and which do not exist in the natural sciences. In short, the methodological distinction between first-level constructs and second-level constructs is a mirror of the distinction between the perspective of the observed people and the perspective of the observing researcher.

A typical and legitimate endeavor in interpretive research is the study of a single setting. For example, an anthropologist can do an ethnography of the Fox Indians in Iowa (Gearing 1970), where the scientific purpose may legitimately be to learn only about the Fox. The anthropologist would face no scientific requirement to produce findings generalizable to other indigenous peoples in North America (although, of course, the same or a different anthropologist could freely choose to do so). Geertz states the following about both theory and generalizability in anthropological studies about culture (Geertz 1973, pp. 25–26): "The essential task of theory building here is not to codify abstract regularities but to make thick description possible; not to generalize across cases but to generalize within them."

Generalizing within a setting stands in contrast to the positivist conception of generalizability, which pertains to generalizing a theory across different settings and where the ideal result would be "universal laws governing every observed phenomenon" (Kolakowski 1968). Where the study of a single setting (e.g., the setting of the Fox in their reservation in Iowa) is an interpretive researcher's objective, generalizing within a setting is not better or worse than, but simply different from, generalizing across settings for a positivist researcher.[7]

There can be many contexts for conceptualizing generalizability. The methodological contrast between positivism and interpretivism is familiar to many IS researchers and provides one good example and insight into the problems typical in developing causal theory. Other examples can be developed. For instance, McKelvey (1982) refers to one kind of science as a "science of uniformity" (p. 12)—i.e., one seeking universal laws. In contrast, McKelvey suggests that there are other forms of theoretical models with similar problems but different essential aims. For example, taxonomic theory is developed for the purposes of classifying entities like animals, plants, or organizations. Similarly phylogenic theory regards the evolution of such entities from their ancestors. McKelvey regards this kind of science as a "science of diversity." Such theories are concerned with identifying commonalities shared by groups of entities, while at the same time establishing differences that distinguish groups. Generalization is engaged by both sciences, one regarding generalized similarities in phenomena, the other regarding generalized differences in phenomena. Taxonomic and phylogenic theories can be developed and applied with contrasting methods concordant with positivist or interpretive frameworks. Examples include numerical taxonomic analysis (generally positivist) or historical analysis (generally interpretive) (McKelvey 1982).

[7] An important point is that the distinction between interpretive research as idiographic and positivist research as nomothetic need not be a rigid one. Consider that generalizing within a case, as in the instance of an ethnography of the Fox, could mean conceptualizing the case to be what a statistically oriented researcher would call the universe or population, across which the interpretive researcher's theory would need to be generalizable. Where the single setting itself is the relevant universe for the researcher, interpretive generalizability and positivist generalizability need not be opposed.

Given the diversity of contexts in which scientists generalize, there is a wide variety of ways of conceptualizing generalizability. In the next section we offer a generalizability framework capable of reaching across diverse and contrasting scientific traditions.

5. A Framework of Four Types of Generalizability

The many ideas about generalizability that we have mentioned suggest a need to bring some order to the diversity of these ideas. A framework could help to organize, identify, and classify different types of generalizability. Such a framework could also offer a clarification to researchers and their critics as to the type(s) of generalizability appropriate and inappropriate to a specific research effort, and for which a researcher should and should not strive. Furthermore, with the statistical, sampling-based notion of generalizability explicitly located in a specific part of a framework, researchers could then readily see statistical generalizability as only a subset of generalizability and as only one among other important notions of generalizability.

One building block in our framework is a distinction implicit in the different notions of generalizability we have examined so far. It is the distinction between empirical statements and theoretical statements. Empirical statements can refer to data, measurements, observations, or descriptions about empirical or real-world phenomena,[8] while theoretical statements posit the existence of entities and relationships that cannot be directly observed, and hence can only be theorized.[9] Both positivist research and interpretive research deal with statements pertaining to the researcher's observations, as well as statements pertaining to the theory that the researcher uses to explain what he or she observes.

Another building block in our framework is another distinction also implicit in the different notions of generalizability we have examined. It is the distinction between what the researcher is generalizing *from* and what the researcher is generalizing *to*. For example, Yin (1984, 1994) mentions generalizing *from* a sample *to* a population, *from* experimental subjects *to* experimental findings, and *from* a case study's findings *to* a theory. Generalizing, according to the definition we cited at the beginning of this essay, refers to generalizing *from* particular instances *to* general notions. Positivism, being nomothetic, requires generalizing *from* a theory *to* different settings.

By joining the two building blocks, we recognize that generalizing can occur in four ways: From empirical statements to other empirical statements, from empirical statements to theoretical statements, from theoretical statements to empirical statements, and from theoretical statements to other theoretical statements. Given the definition of generalize ("to form general notions by abstraction from particular instances"), the four different ways of generalizing indicate that the outputs of generalizing (the "general notions") can be either theoretical statements or empirical statements, and the inputs to generalizing (the "particular instances") can also be either theoretical statements or empirical statements. The result is the framework that appears in Figure 5.

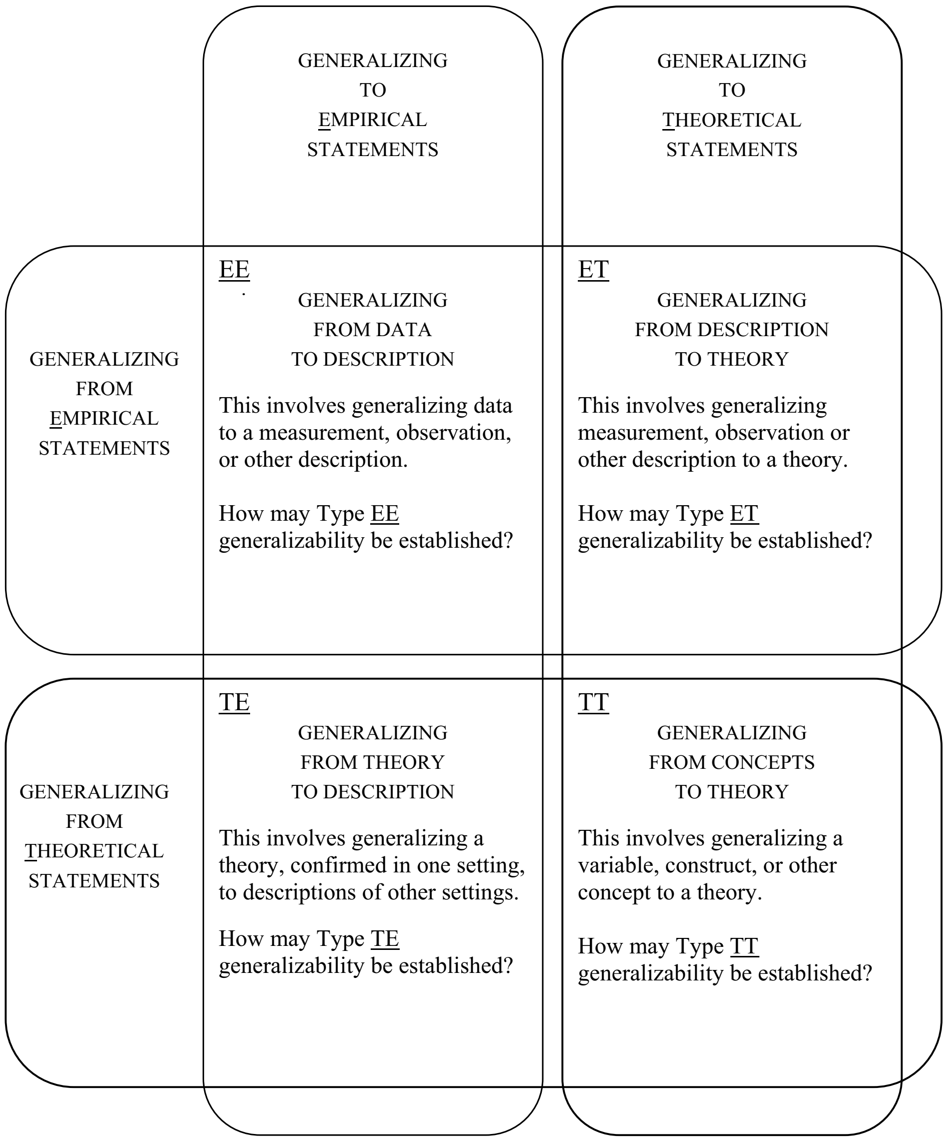
Type EE Generalizability: Generalizing from Data to Description

In Type EE generalizability, the researcher generalizes from empirical statements (as inputs to generalizing)

[8] Examples of empirical statements would be (1) the numerical values of the sample points in statistical sampling, (2) numbers (such as sample estimates) presented as measurements of characteristics of a population, (3) a certain variable's numerical value for an experimental effect that a researcher observes or predicts that she will observe after she administers the experimental treatment, and (4) a case researcher's thick description of the behaviors of the people in a particular organization.

[9] In organizational studies, an example of theoretical statements would be propositions about a particular corporation's culture and social structure, which are not directly observable but whose existence could be theorized from the publicly observable behaviors and actions of the corporation's employees. In physics and astronomy, examples of theoretical statements would be propositions about subatomic particles and black holes, which do not lend themselves to direct observation but whose predicted effects can (if these theorized entities indeed exist) be observed. For the natural and social sciences that express theories mathematically, the theoretical statements could take the form of equations and the empirical statements would include a dataset to which the equations could be applied.

Figure 5 A Generalizability Framework: Four Types of Generalizing and Generalizability



to other empirical statements (as outputs of generalizing). Type EE reasoning involves generalizability in two ways: The generalizability of data to a measurement, observation, or other description (such as a descriptive statistic or a thick description) and the generalizability of the resulting measurement, observation, or other description beyond the sample or domain from which the researcher has actually collected data (such as generalizing to the unsampled portion of the population or to the people in the corporation who were not interviewed). In either case, we regard the product of the generalizing to be a description.

Consider the earlier example about the perceived usefulness, U, that a manager associates with a particular technology, where U is measured on a scale from 1 to 5. Empirical statements describing the data or sample points (i.e., the individual values of U for n different managers in a random sample) could be "$u_1 = 3.10$," "$u_2 = 5.23$," …, "$u_n = 4.91$." These empirical statements would serve as inputs to the process of generalizing, from which the output would be other empirical statements, such as "the sample mean of 2.80 serves as an estimate or measurement of the population mean." The matter of generalizability arises in two ways in this example. First, how generalizable are the sample points to the sample estimate? Second, how generalizable is the sample estimate to the corresponding population characteristic?

The generalizability of sample points to a sample estimate depends on whether certain conditions required by statistical methods are satisfied: In our case, one condition is that the sampling must be done in a properly randomized way, and another condition is that the population being sampled must be normally distributed if the sample size is small and the student's t distribution is to be used instead of the normal distribution for calibrating statistical significance. (Of course, other statistical methods, pertaining to other forms of statistical inference, would pose other conditions to be satisfied.) If such conditions are ignored or otherwise not satisfied, the sample points could not then be properly generalized to a valid sample estimate. However, if such conditions are satisfied, an increase in the sample size can, as previously discussed, lead to an increase in generalizability in the following way: The generalizability of one sample to other samples that could be taken from the same population would increase (i.e., there would be improved generalizability in the sense of reliability) and the generalizability of the sample points to the sample estimate would increase. An improved level of statistical significance (i.e., a smaller p-value) would signify this.

The generalizability of a sample estimate to its corresponding population characteristic is a different matter altogether. Hume's truism, as discussed earlier, prohibits the conclusion that an increase in sample size leads to an increase in the generalizability of any sample estimate to its corresponding population characteristic. To conclude otherwise would require application of the uniformity of nature proposition, the attempted proof of which would trigger the infinite regress in reasoning identified by Hume. Hence, with regard to Type EE generalizability in statistical, sampling-based inference, an increase in the sample size can increase the generalizability of the sample points to a sample estimate, but does not increase the generalizability of the sample estimate to the corresponding population characteristic.

A second example of Type EE generalizability involves the data or empirical statements describing a research subject's responses to the different items in a measurement instrument, such as one for measuring perceived usefulness, U. A research subject's answers (e.g., numerical values on a scale from 1 to 5) to these items would be inputs to the process of generalizing, from which the output would be a generalized empirical statement describing the measurement (e.g., "perceived usefulness = 4.17") for this person. The generalizability of the data to the measurement depends on whether the requirements of instrument-validation procedures have been satisfied. These requirements involve pretest and pilot studies, content and construct validities, and reliability (Boudreau et al. 2001, Smith et al. 1996). In the situation where the measurement instrument has not been validated, the data collected from a research subject would lack generalizability to any valid measurement for that individual.

For a third example of Type EE generalizability, we turn to ethnography. Van Maanen (1983) recounts his participant-observation work with some police officers who made a point of waiting outside a tavern so as to be in a good position to arrest drunk patrons who left the tavern by driving away. The officers explained to Van Maanen that these arrests were necessary for "getting the drunk-hunting sergeant off our backs for awhile" (p. 45); interestingly, however, in their previous conversations with Van Maanen they had presented themselves as autonomous, independent actors. From this particular set of field data, an ethnographer could generalize the description that, in the world of these officers, autonomy is indeed highly valued—so much so, in fact, that the officers will conjure up busywork to satisfy their sergeant, distract his attention, and thereby otherwise preserve their autonomy. Regarding generalizability, the concern is whether the presentational data (what the informants say directly in answer to questions from the ethnographer) and operational data (what the ethnographer observes in the actions and behaviors of the informants) can be generalized into a valid, empirical statement. An example of the ethnographer's resulting empirical statement could be "autonomy is highly valued in the eyes of the officers." Van Maanen states that "the ethnographer must continually assess the believability of talk-based information harvested over the course of a study, an evaluation dependent upon the fieldworker's interest, skill, and good fortune in uncovering lies, areas of ignorance, and the various taken-for-granted features of the studied organization" (pp. 50–51). Data receiving the ethnographer's favorable assessment would be generalizable to a valid descriptive statement (which, in turn, could become part of a larger thick description). In contrast, data receiving no such assessment would lack generalizability to any valid descriptive statement.

The generalizability of data to a descriptive statement is distinct from the generalizability of the descriptive statement to portions of the domain that the ethnographer has not observed. Specifically, in our example, would an increase in the number of officers whom the ethnographer observes serve either to establish or improve the generalizability of the descriptive statement, "autonomy is highly valued in the eyes of the officers," to all those other officers whom the ethnographer has not observed? An affirmative answer would trigger the infinite regress identified in Hume's truism. In other words, no descriptive statement (whether quantitative or qualitative) is generalizable beyond the domain that the researcher has actually observed.

Finally, Yin's concept of Level-1 inference (see Figure 1) also falls under the heading of Type EE generalizability. Level-1 inference can involve using sample estimates (such as the sample mean) as measurements of the corresponding population characteristics (such as the population mean)—where our earlier discussion concluded that a large sample size does not establish the generalizability of sample estimates to population characteristics, but can only establish the generalizability of the sample points to the sample estimate. Level-1 inference can also involve the use of data describing the attributes of research subjects in an experiment as a basis on which to generalize empirical statements about the experimental findings. To the extent that an experiment uses statistical, sampling-based methods where each research subject represents a sample point, our discussion of statistical, sampling-based generalizability applies: Increasing the size of the sample of research subjects does not improve or even help to establish the generalizability of the experimental findings to the population from which the research subjects are sampled, but can only help to improve or establish the generalizability of the data describing the research subjects to the findings of this particular experiment.

All of the preceding examples illustrate how Type EE generalizability recognizes and differentiates the generalizability of the data forming the basis of a measurement, observation, or other description and the generalizability of the resulting measurement, observation, or other description beyond the domain from which the data were collected.

Type EE generalizability nonetheless provides a useful perspective for exposing the very limited realm where statistical, sampling-based generalizability is relevant. Not only is statistical, sampling-based generalizability a subset of Type EE generalizability, but also, Type EE generalizability itself is, in turn, a subset of the overall generalizability framework (as presented in Figure 5), where it is one of four types of generalizability. The framework is therefore useful for, among other things, summarily demonstrating that statistical, sampling-based generalizability is not a general form of generalizability. It is a special case of generalizability. For many IS researchers, such as those identified in Table 1 above, this demonstration has the benefit of liberating them to take advantage of additional forms of generalizability.

Type ET Generalizability: Generalizing from Description to Theory

In Type ET generalizability, the researcher generalizes from empirical statements (as inputs to generalizing) to theoretical statements (as outputs of generalizing). Under this category of generalizability, we may classify certain conceptions of generalizability offered by Yin (1984, 1994); Walsham (1995b); Klein and Myers (1999); Glaser and Strauss (1967); Strauss and Corbin (1998); and Eisenhardt (1989). Analogous to our discussion on Type EE reasoning, we pose that Type ET reasoning can also involve generalizability in two ways: The generalizability of measurements, observations, or other descriptions to theory, and the generalizability of the resulting theory beyond the sample or domain that the researcher observes (such as the unsampled portion of the population or the parts of the organization where the field worker has neither conducted interviews nor collected data in other ways).

Yin (1984, 1994) provides three synonyms for generalizing from empirical to theoretical statements: "Analytical generalization," "Level-2 inference," and "generalizing to theory." Yin offers the following examples of generalizing from empirical to theoretical statements (see Figure 1): Generalizing from experimental findings to theory, generalizing from case study findings to theory, and generalizing from population characteristics to theory. We add the following details to Yin's examples: Empirical descriptions serving as inputs to the process of generalizing could specify, for example, the measurements of the effect of a treatment administered in a particular field experiment; the rich details in a case study of a particular corporate headquarters; or the sample estimates of the population characteristics of workers in a particular geographic region. The resulting theoretical statements could comprise, respectively, a theory positing new variables and the relationships among them that would explain the experimental effect that was measured in the field experiment; a theory explaining the corporate headquarters' social structure and culture that would account for the behaviors and actions noted in the thick description of the case study; or a theory explaining the underlying labor market forces that would result in the levels of the population characteristics that the sample estimated.

Yin focuses on case studies. For empirical descriptions in a case study to be generalizable to a valid theory, Yin prescribes the use of procedures that involve what he calls multiple sources of evidence, a case study database, and member checking—all of which would help to ensure the quality of the descriptions. However, the generalizability of the resulting theory beyond the case actually observed is a different matter. To claim that a theory will remain valid beyond the observed case (i.e., capable of generalizing valid descriptions of field settings not yet observed) would require accepting the uniformity of nature proposition, the validity of which is unestablished and the attempted proof of which would trigger the infinite regress identified in Hume's truism. Hence, a theory generalized from the empirical descriptions in a particular case study has no generalizability beyond the given case. This particular lack of generalizability is not only a feature of qualitative studies, but also statistical, sampling-based studies. Generalizing beyond the given field setting in case research corresponds to generalizing beyond the given population in statistical research. Sample points may be generalized to sample estimates of population characteristics, but certainly have no generalizability beyond the given population.

Although Yin's case research method is considered to be positivist, his concept of analytical generalization has received attention and approval from a prominent interpretive IS researcher, Walsham (1995b). Walsham accepts Yin's notion of generalizing to theory and extends it to four types of generalization. Walsham explains (pp. 70–80) that, beginning with the facts or the rich description of a case, the researcher can generalize to concepts, to a theory, to specific implications, or to rich insight. All four of Walsham's examples involve generalizing from empirical statements (reflecting the observations made in a case study) to theoretical statements (concepts, theory, specific implications, and rich insight).

Klein and Myers (1999) also recognize the process of generalizing from empirical statements to theoretical statements. Whereas they acknowledge that "interpretive research values the documentation of unique circumstances," they also emphasize, "it is important that theoretical abstractions and generalizations should be carefully related to the case study details as they were experienced and/or collected by the researcher . . . ." They add: "The key point here is that theory plays a crucial role in interpretive research, and clearly distinguishes it from just anecdotes" (p. 75). For them, generalizing from idiographic details to theory is so important that they elevate it to one of their seven principles for assessing interpretive field work: The principle of abstraction and generalization.

Long before the work of Yin, of Walsham, and of Klein and Myers, there was the qualitative method that Glaser and Strauss (1967) established, "grounded theory." It embodies the idea that theory is grounded in descriptive categories and relationships that emerge from properly collected and coded data, where the use of theoretical preconceptions or prior theory is minimized so as not to force the emergence of a theory. This squarely fits the phrase "generalizing to theory" and the phrase "generalizing from empirical statements to theoretical statements." Grounded theory has evolved in different directions (Glaser 1992, Strauss and Corbin 1998), where the differences reside in how to operationalize the emergence or grounding of theory, not in the idea that theory emerges from observations.

Eisenhardt (1989) describes how to build theories in case study research, where she explicitly subscribes to Yin's case study method and to Glaser and Strauss's grounded theory. Paré and Elam (1997) operationalize Eisenhardt's theory-building framework for IS research. Eisenhardt's framework for building theories is a framework for generalizing empirical descriptions to theory.

In summary, the notion of the generalizability of empirical descriptions to theory is well developed. Hence, criticisms that case studies and qualitative studies are not generalizable would be incorrectly ruling out the generalizability of empirical descriptions to theory. Furthermore, such criticism could be incorrectly presuming that statistical generalizability is the only form of generalizability.

Type TE Generalizability: Generalizing from Theory to Description

Business-school researchers are interested not only in pure or basic research—the development, testing, and confirmation of theories—but also in the utility of their theories in the actual business settings of executives, managers, consultants, and other practitioners. The particular organization where a practitioner wishes to apply an academically developed theory is likely not to be the same setting where the academic researcher collected her data when developing and testing her theory. Hence, the generalizability of a theory to a description of the results that the practitioner would observe if he were to use the theory in a new setting—i.e., a setting other than the one(s) where the theory was empirically tested and confirmed—is arguably the most important form of generalizability in business-school research. Figure 5 refers to this as Type TE generalizability, which involves generalizing from theoretical statements (in particular, a theory that has already been developed, tested, and confirmed, such as one reported in a published journal article) to empirical statements (here, descriptions of what the practitioner can expect to observe in his specific organization if he were to apply the theory).

Type TE generalizability happens to be closely related to empirical testing in the following way. The empirical testing of a theory can involve applying the theory (as the major premise in a syllogism) to a set of initial conditions (i.e., the minor premise, consisting of empirical statements that describe the conditions observed in the experimental or field setting before the experimental treatment is administered), resulting in the conclusion (i.e., predictions, which are empirical statements describing what should be observed at the end of the experiment if the theory is true). In IS research, the experimental treatment is often the introduction of an information technology (e.g., an enterprise software, a group support system, a CASE technology). However, the validity of the theory in a new setting (e.g., a field setting other than the one[s] where the theory has been empirically tested) would remain an open question. On what basis may a researcher (or a practitioner reading about the researcher's theory in a published journal article) justify the claim that a theory, already empirically tested and confirmed in one setting, is generalizable to the new setting?

The only way in which a researcher (or practitioner) may properly claim that the theory is indeed generalizable to the new setting would be for the theory to be actually tested and confirmed in the new setting. This would involve making a comparison between what the theory would describe as happening in the new setting and what is actually observed as happening in the new setting. This is the procedure advocated by Lee (1989), who attributes it to Campbell (1975).

However, practitioners who read a theory reported in a journal article do not necessarily have the time, resources, or desire to perform a scientific test of the theory in the setting of their own companies prior to their actual application of it. The necessity and urgency of managing real-world problems, hence, suggest a look at the possibility of relaxing the requirements of strict scientific procedures. For instance, this could involve assuming that the uniformity of nature proposition is correct (which, in turn, would obviate the need to retest a previously confirmed theory in a new but similar setting), but making this assumption would mean violating Hume's truism and, hence, also accepting the responsibility for the possibility that the assumption fails. And, whereas this discussion has concentrated on positivist theory, Hume's truism makes no distinction between positivism and interpretivism and, therefore, is no less problematic for an interpretive theory that faces real-world application in a new context (i.e., a setting where it was not developed or tested).[10] An irony is that a scientific researcher's departure from any scientific requirement can give the appearance that university-based scientific researchers, who appeal to the authority of science to legitimize their voice in the practitioner world, are not practicing what they preach. In the section, "Discussion and Illustration," we will return to a discussion of the responsibility that the practitioner and the researcher both need to take in applying theory to new settings—in particular, settings where a theory has not been empirically tested and confirmed.

Type TT Generalizability: Generalizing from Concepts to Theory

In Type TT generalizability, a researcher generalizes from theoretical propositions in the form of concepts (such as a variable, an a priori construct, or another concept) to the theoretical propositions that make up a theory (specifically, a set of logically consistent propositions that, pending the results of empirical testing, could qualify as a theory). DeLone and McLean (1992) draw attention to a problem, specific to IS research, in Type TT generalizability. They observe six very different ways in which different IS research studies have conceptualized the dependent variable, IS success; they are system quality, information quality, information use, user satisfaction, individual impact, and organizational impact. Because of the inconsistent or unreconciled conceptualizations of this dependent variable, one may argue that it is not yet capable of supporting theoretical generalizations across different studies or even within a study. Much the same can be said about information technology in the role of the independent variable in IS theories. The wide-ranging instantiations of information technology (e.g., the Internet, ERP systems, word-processing packages, fax machines, telephones, and even nonelectronic information technologies such as written text) may lead one to question if this variable indeed refers to the same phenomenon across different IS research studies. Because of the varied conceptualizations or underconceptualization of information technology as a variable, one may argue that it too is as yet incapable of supporting theoretical generalizations, either within a study or across different studies.

Another form of generalizing from concepts to theory would be the formulation of a theory based on the synthesis of ideas from a literature review. Such theories can appear in what Zmud (1998) calls "pure theory" manuscripts, where Zmud points to the *Academy of Management Review* for examples. However, there are presently no explicit, general criteria for assessing the capability of variables, constructs, or other concepts to be generalized or otherwise developed into a theory.

6. Discussion and Illustration

A study useful for anchoring a discussion and illustration of the generalizability framework is "Gender Differences in the Perception of E-Mail: An Extension to the Technology Acceptance Model," conducted by

[10] In older disciplines such as anthropology and sociology, which are pure rather than applied sciences, interpretive theories have been crafted to be idiographic rather than nomothetic, so the generalizability issue that Hume's truism raises has not been a concern. However, for business-school disciplines such as information systems, the utility of theories in the management of real-world problems is the raison d'être of much research, whether the theories are positivist or interpretive. This renders Hume's truism both relevant to and problematic for interpretive IS research.

Gefen and Straub (1997). It "extends the TAM model (Davis 1989) and the SPIR addendum (Straub 1994) by adding gender to an IT diffusion model" (p. 389). The resulting theory (henceforth, the extended theory), formulated and tested by Gefen and Straub, has these variables: Perceived usefulness, perceived ease of use, level of system use (from Davis 1989), social presence/information richness or SPIR (from Straub 1994), and gender. The extended theory posits specific direct relationships among these variables. Gefen and Straub's claim is that gender makes a difference in the level of system use. In testing the extended theory with evidence from "only" three firms, the Gefen and Straub study opens itself up to the same criticisms regarding generalizability as did the qualitative case studies cited in Table 1. At the same time, the Gefen and Straub study also uses quantitative data for statistical hypothesis testing. Hence, the Gefen and Straub study is useful for addressing a number of different aspects of generalizability. Using survey evidence on the use of e-mail, Gefen and Straub conclude that gender does make a difference to system use.

Because of the importance that *Type TE generalizability* (which involves *generalizing from theory to description*) has for real-world applications of theory, we will emphasize this type of generalizability in our discussion of Gefen's and Straub's (1997) example. This refers to the capability to generalize Gefen's and Straub's extended theory to descriptions of what would happen in settings other than the three firms that they observed. On this note, Gefen and Straub themselves are not encouraging: "From the standpoint of external validity, the study gathered data from three firms in one industry across three countries, which, *per force*, limits the generality of the results." Their logic is that if they had sufficiently increased the number of firms[11] from which they collected their data, then the problem of the limited "generality of the results" could have been resolved. There are two caveats that we add to their logic.

First, even if Gefen and Straub (1997) had empirically tested and confirmed their extended theory in an overwhelming number of different firms, one would be able to claim only that the theory is generalizable to these firms and no others. This is a consequence of Hume's truism. For scientific researchers to go beyond this claim—for example, by applying the unfounded uniformity of nature proposition to assert that the theory is generalizable to other organizational settings that are similar to any of the firms where the theory was confirmed—would require the researchers to depart from the authority of science as the basis for their expertise and the basis for action. It would be the ethical responsibility of editors and authors to make this caveat explicit to practitioners, who are the intended recipients of the research. At the same time, it would be ironic for scientific researchers to violate their allegiance to science when recommending their research to practitioners. This is a dilemma that cannot be resolved in this essay, but action research suggests how it could be managed—a matter to which we will return at the end of this section.

The second caveat is that Gefen and Straub (1997) would need to redo their empirical testing so that it examines just one firm at a time. In their study, they tested their hypotheses with data combined across the three organizations. The combining of the data, however, makes it possible for the extended theory to survive hypothesis testing in the situation where, for instance, the extended theory is false in one setting but true in the two others, with the evidence in the latter two overwhelming the evidence in the first. To make the claim that a theory can be generalized to a particular setting requires that it be empirically tested in that setting. For Gefen and Straub's study, this would mean breaking up their overall sample into three parts, where each part contains sample points collected from just one of the three firms. This would necessarily result in three samples sizes, each one of which is smaller than the total sample size that Gefen and Straub used in their study. For Gefen and Straub, this would lead to the additional complication in which the smaller sample sizes would have the unfavorable effect of diminishing the generalizability of the sample points to the sample estimate (i.e., worsened levels of statistical significance), which is an instance of the larger category of *Type EE generalizability* and which involves the *generalizability of data to measurement, observation, or other descriptions.* Statistical

[11] One may say the same about the number of industries and the number of countries.

methods themselves do not prohibit the combining of data across organizational settings; however, the added requirement of establishing Type TE generalizability requires that empirical testing be conducted within, not across, individual organizational settings. This in turn would require, for each organizational setting, a sample size sufficiently large to allow for statistical significance. Note, however, that neither increasing the sample size in each organizational setting nor increasing the number of such organizational settings may ever overcome Hume's truism. *A theory may never be scientifically generalized to a setting where it has not yet been empirically tested and confirmed*.

Though Gefen's and Straub's (1997) study is positivist, two major points in our discussion so far also apply to interpretive research. First, a necessary step in making a theory generalizable to different organizational settings is that the theory—whether positivist or interpretive—be examined in each of these organizational settings separately. Second, the dilemma between scientific validity (complying with Hume's truism) and practical impact (applying a theory in a new organizational setting) makes no distinction between positivist theory and interpretive theory.

A third major point is that statistical, sampling-based generalizability remains as valid a concept as ever. However, the Gefen and Straub example illustrates the limits of the domain of this concept: It is relevant when considering Type EE generalizability, not TE generalizability.

A few more remarks about Type EE generalizability are in order. First, Gefen and Straub (1997) do not report how random or representative their sample is. Hence, the generalizability of their sample points to sample estimates (which are "betas" measured through the statistical technique of PLS) is unknown. Second, if their sample is indeed representative, then one may claim that the sample points are generalizable to the sample estimates in three of the four cases of Gefen's and Straub's hypothesis testing; in these three cases, the levels of statistical significance were 0.05 or better. Third, because Gefen and Straub were careful to use validated instruments in their survey, the data that they collected from each survey respondent were generalizable to valid measurements for that individual.

Also requiring discussion are *Type ET generalizability* and *Type TT generalizability*, the explicit consideration of which could benefit Gefen's and Straub's (1997) extended theory. There are a number of ways in which Type ET generalizability—which involves *generalizing from empirical statements to theoretical statements*—can be relevant to Gefen's and Straub's research. First, one might suggest that they could have conducted a field study (resulting in rich empirical statements) so as to formulate a better model (better theoretical statements, resulting from the rich empirical statements) prior to their quantitative data collection and statistical testing. Such a suggestion, however, would run counter to established positivist conventions and be inappropriate and simply unfair—as would be a suggestion that an interpretive case study be reframed as a statistical experiment. A better suggestion would be for a follow-up case study that focuses on how and why Gefen's and Straub's fourth hypothesis ("Women's use of e-mail will be greater than that of men") received no empirical support. If the case study is performed in a recognized manner (e.g., according to Yin's case study procedures), then the resulting empirical statements could be considered valid, and hence constitute material from which to generalize theoretical statements that embody a refinement of Gefen and Straub's original extended theory. The validity and generalizability of the new theory would, of course, then need to be established (perhaps in yet another study, which could be statistical) through empirical testing.

Type TT generalizability, which involves *generalizing a variable, construct, or other concept to a theory*, becomes relevant to Gefen and Straub (1997) in their generalizing of different concepts taken from the research literature (concepts such as gender, culture, and perceived usefulness) to their extended model. One might question how ready these concepts are for being generalized to a new theory; however, research methodologists have not yet provided a way to answer such a question.

Finally, what might scientific researchers do so that they may recommend their theories for application in new settings, where the theories have not yet been empirically tested and confirmed in those settings? One possibility would be for the researchers who make, and the practitioners who accept, such recommendations to share responsibility for any unfavorable outcome. Providing a model for this would be action research, where practitioners and scientific researchers work as a team, share responsibility, regard the theory as tentative, apply it, and then improve it over successive cycles of application and reflection until the practitioner-defined problem is adequately addressed. Examples of action research in information systems are Straub and Welke (1998) and Baskerville and Stage (1996). The process of action research suggests that the ability of a theory to be generalized to a new setting could also depend on factors outside the theory itself: The successful forming of a researcher-practitioner team; the sharing of responsibility by the team members; the ability of the researchers to express the scientific theory in practitioner-understandable ways; the ability of the practitioners to express the business problem in researcher-understandable ways; the sharing of the understanding that their application of the theory serves both as an empirical test of it and as a means of solving the practitioner-defined problem; and the sharing of the attitude that they have the purpose of learning from the application-reflection cycles and continually refining the theory, at least until the practitioner problem is solved. In this manner of using theory in practice, there is a social process for testing, refining, and hence circumspectly generalizing the theory to a setting where it was not previously developed or tested. For the future, action research teams might consider additionally reflecting on their experience in generalizing a scientific theory to descriptions (i.e., descriptions of what the action research team expects to observe in their organization upon applying the theory) so as to contribute to the eventual development of guidelines that other action research teams, or even practitioners alone, could use for applying theory in new settings. The development and the "beta testing" of such guidelines might even justify a new, major research stream.

7. Conclusion

In a case study, the researcher may appropriately strive to develop a theory that is generalizable within the case setting. In a statistical study, the researcher may appropriately test a theory with the help of measurements of population characteristics through sample points that are generalizable to sample estimates. In neither case, however, would it be appropriate to criticize a theory for a lack of generalizability to other settings; the reason is that, as a consequence of Hume's truism, a theory may *never* be generalized to a setting where it has not yet been empirically tested and confirmed. Along the same lines, neither an increase in the sample size in a statistical study nor an increase in the number of sites in a multisite case study would be an indicator of greater generalizability of a theory to new settings. And the suggestion that a theory developed in a qualitative case study can achieve greater generalizability to new settings through an increase in its sample size would involve the double error of violating Hume's truism and conflating Type TE generalizability and Type EE generalizability. Whether research is conducted quantitatively or qualitatively, there is only one scientifically acceptable way to establish a theory's generalizability to a new setting: It is for the theory to survive an empirical test in that setting.

Researchers have been using (or, in effect, generalizing) the term "generalizability" to refer to many different concepts, some of which go beyond the *Oxford English Dictionary*'s definition and have caused considerable confusion. Consider that the type of generalizability that case studies have been incorrectly singled out for lacking—Type TE generalizability, which involves generalizing a theory confirmed in one setting to descriptions of other settings—refers to reasoning from theoretical statements to empirical statements, which is actually deduction, not induction. Indeed, such reasoning is the opposite of the *OED* definition of generalize, which is "to form general notions by abstraction from particular instances." In this light, if there is a quality of case studies that might merit criticism, it would be a lack of "particularizability," as it were, rather than generalizability. Our generalizability framework is an attempt to bring some clarity to the many different concepts in the scholarly discourse on generalizability. Additional methodological research on generalizability itself can be helpful in advancing the development of scientific theories, particularly the development of theories relevant in real-world settings.

Acknowledgments

The authors thank the three reviewers and associate editor for provocative insights that shaped this essay. The authors are also grateful for the feedback on earlier versions of this essay presented at research seminars at the City University of Hong Kong, Colorado State University, George Washington University, Georgia State University, Indiana University at Bloomington, the London School of Economics and Political Science, the Queen's University of Belfast, the Stockholm School of Economics, Temple University, Texas Tech University, the University of Nebraska at Lincoln and at Omaha, the Doctoral Consortium of the 2001 European Conference on Information Systems, and the Seventh Annual (2001) Summer School in Design and Management of Information Technology at the Magleås Kursuscenter in Denmark. This essay began as "Distinctions among Different Types of Generalizing in Information Systems Research," which the authors presented at the 1999 meeting of Working Group 8.2 of the International Federation for Information Processing.

References

Baskerville, R., J. Stage. 1996. Controlling prototype development through risk analysis. *MIS Quart.* **20**(4) 481–504.

Berger, P., T. Luckmann. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge.* Anchor Press, New York.

Boland, R. 1985. Phenomenology: A preferred approach to research on information systems. E. Mumford, R. Hirschheim, G. Fitzgerald, T. Wood-Harper, eds. *Research Methods in Information Systems*. Elsevier Science Publishers, Amsterdam, The Netherlands, 193–201.

Boudreau, M., D. Gefen, D. Straub. 2001. Validation in information systems research: A state-of-the-art assessment. *MIS Quart.* **25**(1) 1–16.

Burrell, G., G. Morgan. 1979. *Sociological Paradigms and Organisational Analysis*. Heinemann, London, U.K.

Campbell, D. 1975. "Degrees of freedom" and the case study. *Comparative Political Stud.* **8**(2) 178–193.

——, J. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin, Boston, MA.

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

DeLone, W., E. McLean. 1992. Information systems success: The quest for the dependent variable. *Inform. Systems Res.* **3**(1) 60–95.

Eisenhardt, K. 1989. Building theories from case study research. *Acad. Management Rev.* **14**(4) 532–550.

Feigl, H. 2002. *Positivism*. Encyclopaedia Britannica. Retrieved May 9, 2002, from http://search.eb.com/eb/rticle?eu=115436.

Garfinkel, H. 1967. *Studies in Ethnomethodology*. Prentice-Hall, Englewood Cliffs, NJ.

Gearing, F. 1970. *The Face of the Fox*. Aldine Publishing, Chicago, IL.

Geertz, C. 1973. *The Interpretation of Cultures*. Basic Books, New York.

Gefen, D., D. Straub. 1997. Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quart.* **21**(4) 389–400.

Glaser, B. G. 1992. *Emergence vs. Forcing: Basics of Grounded Theory Analysis*. Sociology Press, Mill Valley, CA.

——, A. L. Strauss. 1967. *The Discovery of Grounded Theory, Strategies for Qualitative Research*. Aldine Publishers, Chicago, IL.

Habermas, J. 1988. *On the Logic of the Social Sciences*. S. W. Nicholsen, J. A. Stark, trans. MIT Press, Cambridge, MA.

Hirschheim, R., H. Klein. 1989. Four paradigms of information systems development. *Comm. ACM.* **32**(10) 1199–1216.

Klein, H. K., M. D. Myers. 1999. A set of principles for conducting and evaluating interpetive field studies in information systems. *MIS Quart.* **23**(1) 67–93.

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

Lee, A. S. 1989. Case studies as natural experiments. *Human Relations* **42**(2) 117–137.

——. 1991. Integrating positivist and interpretive approaches to organizational research. *Organ. Sci.* **2**(4) 342–365.

——. 1999. Researching MIS. W. Currie, R. Galliers, eds. *Rethinking MIS.* Oxford University Press, New York, 7–27.

——, J. Liebenau, J. DeGross, eds. 1997. *Information Systems and Qualitative Research*. Chapman & Hall, London, U.K.

Luthans, F., T. Davis. 1982. An idiographic approach to organizational behavior research: The use of single case experimental designs and direct measures. *Acad. Management Rev.* **7**(3) 380–391.

McKelvey, B. 1982. *Organizational Systematics: Taxonomy, Evolution, Classification*. University of California Press, Berkeley, CA.

Mumford, E., R. Hirschheim, G. Fitzgerald, T. Wood-Harper, eds. 1985. *Research Methods in Information Systems*. Elsevier Science Publishers, Amsterdam, The Netherlands.

Nissen, H.-E., H. K. Klein, R. A. Hirschheim, eds. 1991. *Information Systems Research: Contemporary Approaches & Emergent Traditions*. Elsevier Science Publishers, Amsterdam, The Netherlands.

Orlikowski, W., J. Baroudi. 1991. Studying information technology in organizations: Research approaches and assumptions. *Inform. Systems Res.* **2**(1) 1–28.

Oxford English Dictionary. 1989. Oxford University Press. Retrieved July 5, 2002, from http://www.viva.lib.va.us/viva/collect/books.html.

Paré, G., J. J. Elam. 1997. Using case study research to build theories of information technology implementation. A. S. Lee, J. Liebenau, J. I. DeGross, eds. *Information Systems and Qualitative Research*. Chapman & Hall, New York, 31–68.

Rosenberg, A. 1993. Hume and the philosophy of science. D. Norton, ed. *The Cambridge Companion to Hume.* Cambridge University Press, New York, 64–89.

Sanday, P. R. 1983. The ethnographic paradigm(s). J. Van Maanen, ed. *Qualitative Methodology*. Sage, Beverly Hills, CA.

Schutz, A. 1962–1966. Concept and theory formation in the social sciences, *Collected Papers*. M. Nijhoff, The Hague, The Netherlands, 48–66.

Smith, H., S. Milberg, S. Burke. 1996. Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quart.* **20**(2) 167–196.

Straub, D. 1994. The effect of culture on IT diffusion: E-mail and FAX in Japan and the U.S. *Inform. Systems Res.* **5**(1) 23–47.

——, R. J. Welke. 1998. Coping with systems risk: Security planning models for management decision-making. *MIS Quart.* **22**(4) 441–469.

Strauss, A., J. Corbin. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 2nd ed. Sage, Thousand Oaks, CA.

Trauth, E., ed. 2001. *Qualitative Research in Information Systems: Issues and Trends*. Idea Group Publishers, Hershey, PA.

Van Maanen, J. 1983. The fact of fiction in organizational ethnography. J. Van Maanen, ed. *Qualitative Methodology*. Sage, Beverly Hills, CA.

Walsham, G. 1993. *Interpreting Information Systems in Organizations*. John Wiley and Sons, New York.

——. 1995a. The emergence of interpretivism in IS research. *Inform. Systems Res.* **6**(4) 376–394.

——. 1995b. Interpretive case studies in IS research: Nature and method. *Eur. J. Inform. Systems* **4**(2) 74–81.

Wood, A. 2000. *Hume: The Problem of Induction*. Stanford University. Retrieved August 31, 2002, from http://www.stanford.edu/~allenw/Phil102/Hume%20-%20Induction.doc.

Yin, R. 1984. *Case Study Research: Design and Methods*. Sage Publications, Beverly Hills, CA.

——. 1994. *Case Study Research: Design and Methods*, 2nd ed. Sage Publications, Thousand Oaks, CA.

Zmud, R. 1998. Editor's comments. *MIS Quart.* **22**(2) xxxix–xxxii.

Lynne Markus, Associate Editor. This paper was received on November 19, 1999, and was with the authors 28 months for 2 revisions.