

GENERALIZING GENERALIZABILITY IN INFORMATION SYSTEMS RESEARCH

Version of 19 July 2001

This paper is a revision of “Distinctions among Different Types of Generalizing in Information Systems Research,” which was submitted to *Information Systems Research* (log #1999-120-000).

Allen S. Lee
Virginia Commonwealth University
AllenSLee@alum.mit.edu

Richard Baskerville
Georgia State University
baskerville@acm.org

The authors presented the first version of the research reported in this paper at the meeting of IFIP Working Group 8.2, held in St. Louis, in August 1999. Based on feedback at that meet, the authors revised the paper and submitted it to *Information Systems Research*, which offered a “revision” decision in May 2000. In the spring of 2001, the first author presented a major revision of this paper in research seminars at the City University of Hong Kong; George Washington University; the Seventh Annual Scandinavian PhD Summer School in the Design and Management of Information Technology; Queen’s University of Belfast; and the London School of Economics. The authors are grateful for the feedback and suggestions received throughout this process.

Generalizing Generalizability in Information Systems Research

Abstract

The concept of generalizability is not homogeneous and monolithic, but can be analyzed into four types: the generalizability of a theory to different settings, the generalizability of a theory within a setting, the generalizability of a measurement or observation, and the generalizability of a variable, construct, or other concept. In this study, we affirm the legitimacy of the statistical, sampling-based conception of generalizability (which falls under the generalizability of a measurement or observation), where we identify it as but a subset of the overall concept of generalizability. We explain that to regard it as the general form of generalizability would, in itself, be an overgeneralization. We offer constructive guidelines, based on established arguments in the philosophy and methodology of science, for how to achieve the generalizability of a theory to different settings. We provide a concrete example of the guidelines by applying them to a classic study (Markus, 1983) to show how a theory may be properly generalized to different settings.

I. Introduction

Quantitative and qualitative research methods in the academic information systems (IS) discipline have experienced significant advances in the past decade. IS researchers have innovated methods or adapted them from other disciplines and then applied these methods in their research. One remaining problematic area, however, pertains to the methodological notion of generalizability. Despite advances that methodologists have made regarding the notion of generalizability, many IS researchers have not availed themselves of these advances. Instead, they have restricted themselves to just one particular notion of generalizability – namely, a statistical, sampling-based notion of generalizability – and applied this notion even outside the bounds of statistical, sampling-based research. As a result, these investigators have forgone claims to generalizability when, in fact, they have not yet explicitly considered conceptions of generalizability appropriate to their research. Our objective in this paper is a coherent presentation of generalizability that makes the contribution of clarifying the methodological conditions under which IS researchers may strive for and achieve generalizability, where these conditions do not displace, but instead, include and add to those of the well known and widely accepted conditions associated with statistical, sampling-based research.

We use the term “generalizability” to refer to the capability of research findings to be valid beyond the immediate research setting (e.g., the particular corporation in a case study, the particular laboratory setting in an experiment, the particular sample in a statistical study) where the findings are established. We use the adjective “generalizable” to describe research findings that can be transferred to and remain valid in a setting outside the original research setting where they were established; the verb, “generalize,” to refer to the actions by which researchers attempt to achieve generalizable results, (e.g., to generalize from a sample to a population); the gerund, “generalizing,” to name the process by which researchers generalize (e.g., the process of generalizing from a sample to a population); and the noun, “generalization,” to refer either to the product of an attempt to generalize (e.g., a statistical generalization) or, as a synonym for generalizing, to refer to the process by which researchers generalize (e.g., “analytic generalization” and “statistical, sampling-based generalization”).

Figure 1, below, offers a preview of each of the different forms of generalizability that this paper will examine. Statistical generalizability (involving the use of a random sample to

describe population characteristics), as this paper will explain, falls in the cell, “generalizability of a measurement or observation,” in Figure 1 and therefore constitutes just a subset of the

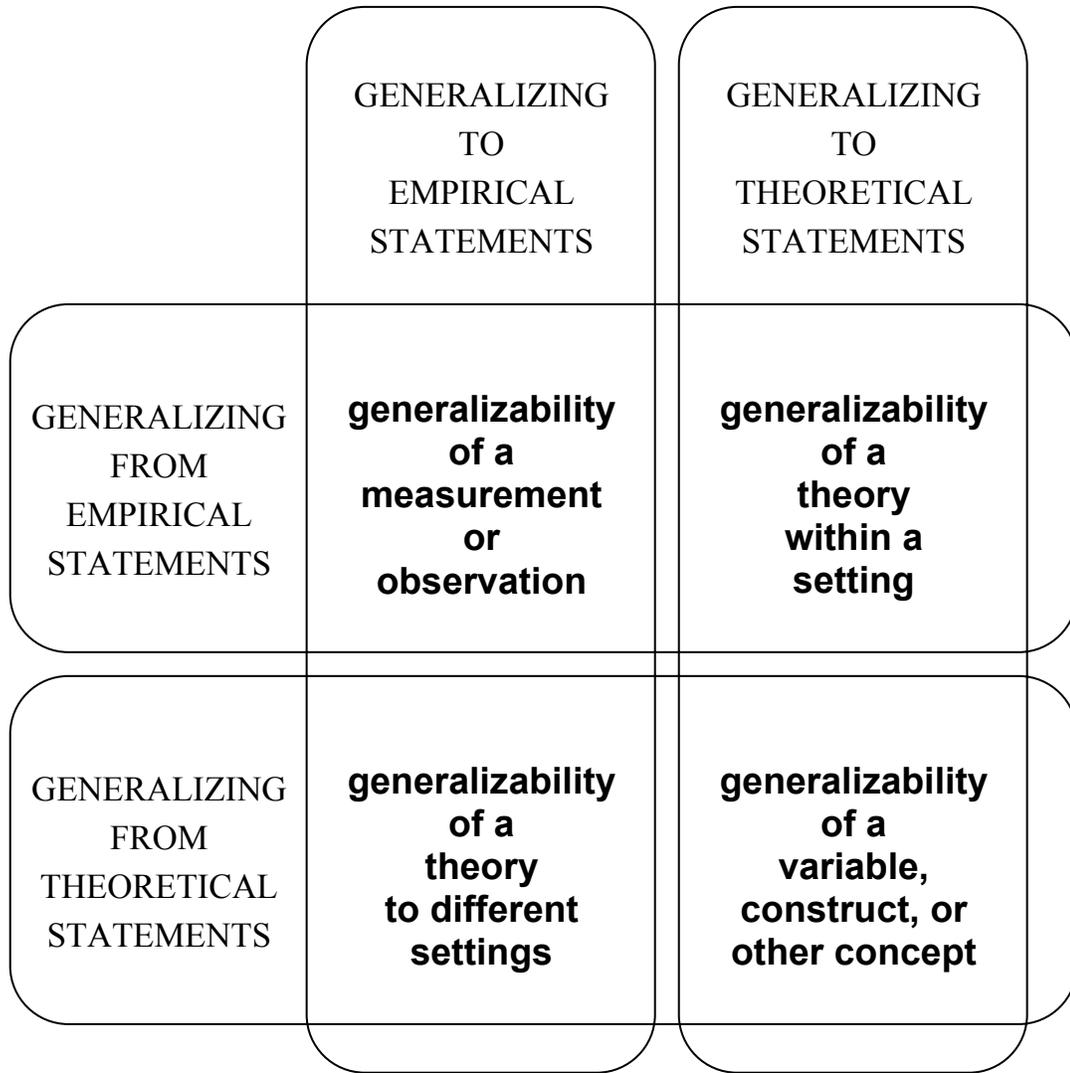


Figure 1

overall universe of generalizability, rather than the general case of generalizability.

In order to show the extent to which the notion of “generalizability” has been misconstrued or underdeveloped in the practice of research, the next part (section II) of the paper will present examples of published IS research where the investigators have applied a statistical, sampling-based notion of generalizability when doing non-statistical, non-sampling research. In section III, we will review four forms of generalizing and four forms of generalizability, all of

which we will arrange in a single map in Figure 3. We reserve our critique for section IV, where we will draw attention to statistical, sampling-based generalizability. In section IV, we emphasize that statistical, sampling-based generalizability is a valid and well respected research notion, but also that philosophers and methodologists have long known about a particular way in which the idea of a greater sample size or number of observations can be misapplied with respect to generalizing the validity of a theory. In section V, based on the map of different forms of generalizing and generalizability from section III and the critique of statistical generalizability in section IV, we will offer guidelines for how researchers may, and may not, strive for and achieve generalizability – in particular, the generalizability of a theory to different settings.

II. Evidence of the Application of Statistical Generalizability in Non-Statistical Research Contexts

In his classic book on case study research, Yin (1994) examines generalization and identifies one particular form of it with the name, “statistical generalization” (p. 38). Yin’s naming of this conception is noteworthy because it signifies that the single word “generalization” itself is not synonymous with statistical generalization, but needs to be qualified in order to refer to generalization in this particular form. In this paper, we will refer to it as “statistical, sampling-based generalization” in order to emphasize that it conceives of generalization not only quantitatively, but also in terms of sampling-based inference. Specifically, it refers to attempts to generalize from a random sample to the population from which the sample is taken. In statistical, sampling-based generalization, a larger sample size is preferred to a smaller one.

Although generalizability is itself a rich and multifaceted concept, IS researchers have largely restricted themselves to just a statistical, sampling-based conception of generalizability. Applying this conception is appropriate for situations involving statistical inference – i.e., situations involving the attempt to measure population characteristics through a sample taken from the population. What is surprising, however, is that IS researchers have applied this notion even outside the bounds of statistical, sampling-based research. It is as if statistical, sampling-based generalizability has itself been overgeneralized, as it were, to non-statistical, non-sampling forms of research. In this section of the paper, we provide examples of published IS research to

indicate the widespread extent to which IS researchers have limited themselves to a statistical, sampling-based notion of generalizability even when they are not doing any statistical sampling.

The principle of generalizing from a sample to a population is firmly established and widely accepted; however, we also emphasize that this principle pertains only in the realm of statistical inference. Researchers who extend the notion of statistical, sampling-based generalization to other forms of empirical scientific research (such as qualitative case studies) would, of course, themselves be making a generalization requiring justification.

As we will show in the next section of this paper, methodologists have offered additional conceptions of generalization that go beyond statistical, sampling-based generalization. One might suppose that IS researchers trained in doctoral programs would be aware of the variety of different conceptions of generalization that now exist and would, therefore, not summarily impose a statistical, sampling-based conception of generalization on non-statistical, non-sampling research. However, as the examples in Table 1 will make evident, the statistical, sampling-based conception of generalizability is widely used in non-statistical, non-sampling research that even well regarded IS researchers have published and presented in top IS journals and conferences. Applying statistical, sampling-based generalization as the norm for what generalization is, these researchers express concerns about what they believe to be the limited generalizability of their non-statistical, non-sampling research. These are concerns that, as we will explain later in this paper, are misplaced because the statistical conception of generalizability is not applicable or appropriate to these researchers' non-statistical circumstances.

In Table 1, we draw attention to particular words that indicate a statistical, sampling-based conception of generalization. For instance, if a case study does not perform statistical inference but nonetheless uses the word "sample," attributes limited generalizability to a small number (e.g., "only one organization," "only two cases," "a single site," "a small sample size"), or otherwise indicates statistical reasoning as its methodological norm-in-use, we underline the signifying word or phrase. The picture that emerges is that the ruling norm of generalizability in these studies is statistical, sampling-based generalizability.

Table 1:
Published Research that Applies the Statistical, Sampling-based Conception of
Generalizability to Non-Statistical, Non-Sampling Research:
Examples from Case Research

Article	Quotation from Article
<p>Majchrzak, A., Rice, R., Malhotra, A., King, N., and Sulin, B. (2000) "Technology Adaptation: the Case of a Computer-Supported Inter-organizational Virtual Team." <i>MIS Quarterly</i>, 24 (4), 569-600. Winner of the award for <i>MIS Quarterly</i> Best Paper of the Year 2000.</p>	<p>There are, of course, many methodological limitations of this study. As a <u>small sample-size, single case study</u>, generalizability cannot be assessed...</p>
<p>Tyran, C., Dennis, A., and Vogel, D. (1992) "The Application of Electronic Meeting Technology to Support." <i>MIS Quarterly</i>, 16 (3), 313-334.</p>	<p>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 <u>statistical</u> analysis and causal inference. Finally, the <u>sample</u> was relatively <u>small</u>, involving <u>five</u> organizations and <u>eight</u> cases.</p>
<p>Brown, C. (1997) "Examining the Emergence of Hybrid IS Governance Solutions: Evidence from a Single Case Site." <i>Information Systems Research</i>, 8 (1), 69-94.</p>	<p>First and foremost, it should be reaffirmed that the <u>single case</u> research strategy employed here <u>only allows</u> generalizability to a research model, which in turn needs to be tested under a multiple case study design or by other field methods.</p>
<p>Vandenbosch, B. and Ginzberg, M., (1996) "Lotus Notes and Collaboration: Plus Ça</p>	<p>We <u>must not generalize</u> from a <u>single case study</u> in one organization.</p>

<p>Change....” <i>Journal of Management Information Systems</i>, 13 (3), 65-81.</p>	
<p>Article</p> <p>Stuchfield, N, and Weber, B. (1992) “Modeling the Profitability of Customer Relationships: Development and Impact of Barclays de Zoete Wedd’s BEATRICE.” <i>Journal of Management Information Systems</i>, 9 (2), 53-76.</p>	<p>Quotation from Article</p> <p>While a <u>single-site</u> case study <u>limits the ability to generalize</u>, there is support for the result that a well-executed information system based on ABC principles can improve management decision making and organizational performance.</p>
<p>Article</p> <p>Newman, M. and Sabherwal, R. (1996) “Determinants of Commitment to Information Systems Development: A Longitudinal Investigation.” <i>MIS Quarterly</i>, 20 (1), 23-54.</p>	<p>Quotation from Article</p> <p>The findings of this paper may also be <u>limited</u> because the paper is based on <u>only one case study</u>. 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.</p>
<p>Article</p> <p>Albers, M., Agarwal, R., and Tanniru, M. (1994) “The Practice of Business Process Reengineering: Radical Planning and Incremental Implementation in an IS Organization” in <i>SIGCPR '94. Proceedings of the 1994 Computer Personnel Research Conference on Reinventing IS: Managing Information Technology in Changing Organizations</i>, ACM Press, Arlington, VA, 87-96.</p>	<p>Quotation from Article</p> <p>From the point of view of theory development, while case studies provide useful anecdotal information, the <u>generalizability from one specific instance to another is often limited</u>.</p>
<p>Article</p> <p>Nandhakumar, J., and Jones, M. (1997) “Designing in the Dark: the Changing User-Developer Relationship in Information Systems Development” in <i>Proceedings of the Eighteenth International Conference on Information Systems</i> (Kumar, K. and DeGross, J., eds.),</p>	<p>Quotation from Article</p> <p>From the evidence of the <u>two cases</u>, it was <u>not possible to identify any generalisable strategies</u> 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</p>

<p>75-87. Winner of the 1997 ICIS Best Paper Award.</p>	<p>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.</p>
<p>Article</p> <p>Robey, D., and Sahay, S. (1996) "Transforming Work through Information Technology: A Comparative Case Study of Geographic Information Systems in County Government." <i>Information Systems Research</i>, 7 (1), 93-110.</p>	<p>Quotation from Article</p> <p>Because they are drawn from a study of <u>two organizations</u>, these results <u>should not be generalized</u> 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.</p>
<p>Article</p> <p>Hidding, G. (1998) "Adoption of IS Development Methods across Cultural Boundaries" in <i>Proceedings of the Nineteenth International Conference on Information Systems</i> (Hirschheim, R., Newman, M., and DeGross, J., eds.), 308-312.</p>	<p>Quotation from Article</p> <p>In particular, in-depth analysis of extensive data <u>from only one organization reduces generalizability</u>, but increases correspondence to reality.</p>
<p>Article</p> <p>Jarvenpaa, S. L., and Leidner, D. (1997) "An Information Company in Mexico: Extending the Resource-Based View of the Firm" in <i>Proceedings of the Eighteenth International Conference on Information Systems</i>, (Kumar, K. and DeGross, J., eds.), 75-87.</p>	<p>Quotation from Article</p> <p>The study has a number limitations that need to be considered in making any conclusions. First, <u>the single case site limits the generalizability of results</u>. 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).</p>

Article	Quotation from Article
Stein, E., and Vandenbosch, B. (1996) “Organizational Learning During Advanced System Development: Opportunities and Obstacles.” <i>Journal of Management Information Systems</i> , 13 (2), 115-136.	Conducting additional case studies will provide instances of the various learning/outcome combinations, and we encourage such research. On the other hand, <u>case studies alone will not result in validity or generalizability.</u> 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.

The preceding examples show that there have been instances where IS researchers apply a statistical, sampling-based conception of generalizability even when they are doing non-statistical, non-sampling research.

The examples in the next table offer a variation on this theme. Unlike the studies in Table 1, the studies in Table 2 involve statistical sampling within a case. By “a case,” we are referring to a company, an organization, a firm, an industry, a technology, or a country. For instance, the researcher takes a sample within a company or does a survey involving one technology. In this situation, the sample size within a case is sufficiently large from the perspective of statistical inference, but the number of cases is small, with the result that the researchers conclude that their research lacks generalizability. As these examples show, these researchers are applying the conception of sampling-based generalization in a research context that cuts *across* cases and are therefore concluding that generalizability would require an increase in the number of cases (e.g., the number of technologies studied, the number of industries where the survey is administered) to a level that would bring about statistical significance – which is a remedy that would be often be impractical (a study of 30 technologies?) and that would not take advantage of non-statistical conceptualizations of generalizability. As in the preceding examples, we underline words or phrases that signify use of the statistical, sampling-based conception of generalizability.

Table 2:
Published Research Whose Authors Perceive No Problem with Sample Size Within a Case, but a Problem with Sample Size Across Cases

Article	Quotation from Article
Gefen, D. and Straub, D. (1997) "Gender Differences in the Perception and Use of E-mail: an Extension to the Technology Acceptance Model." <i>MIS Quarterly</i> , 21 (4), 389-400.	From the standpoint of external validity, the study gathered data from <u>three firms</u> in <u>one industry</u> across <u>three countries</u> , which, per force, <u>limits the generality</u> of the results.
Researchers and Journal	Quotation
McKeen, J. and Guimaraes, T. (1997) "Successful Strategies for User Participation in Systems Development." <i>Journal of Management Information Systems</i> , 14 (2), 133-150.	In addition, we <u>sampled from only eight organizations</u> - <u>not a large sample</u> - and attempts to generalize our results to the larger <u>population</u> of organizations may not be warranted. Because of these methodological limitations, readers should be careful when applying these results to their organizations.
Researchers and Journal	Quotation
Goodhue, D. and Thompson, R. (1995) "Task-Technology Fit and Individual Performance." <i>MIS Quarterly</i> , 19 (2), 213-226.	Though we need to be careful about generalizing too freely about the impact of specific factors of TTF from a <u>sample including only two companies</u> (including more companies in our <u>sample</u> might bring other factors into sharper focus), the results do strongly support Proposition 3.
Researchers and Journal	Quotation
Mukhopadhyay, T., Kekre, S. and Kalathur, S. (1995) "Business Value of Information Technology: a Study of Electronic Data Interchange." <i>MIS Quarterly</i> , 19 (2), 137-156.	Our results underscore the potential benefit and value of EDI and to some extent, IT in general. Although <u>no attempt should be made to directly generalize</u> the gains [at Chrysler Corporation] to other sites or other technologies, the lessons from this study are significant.

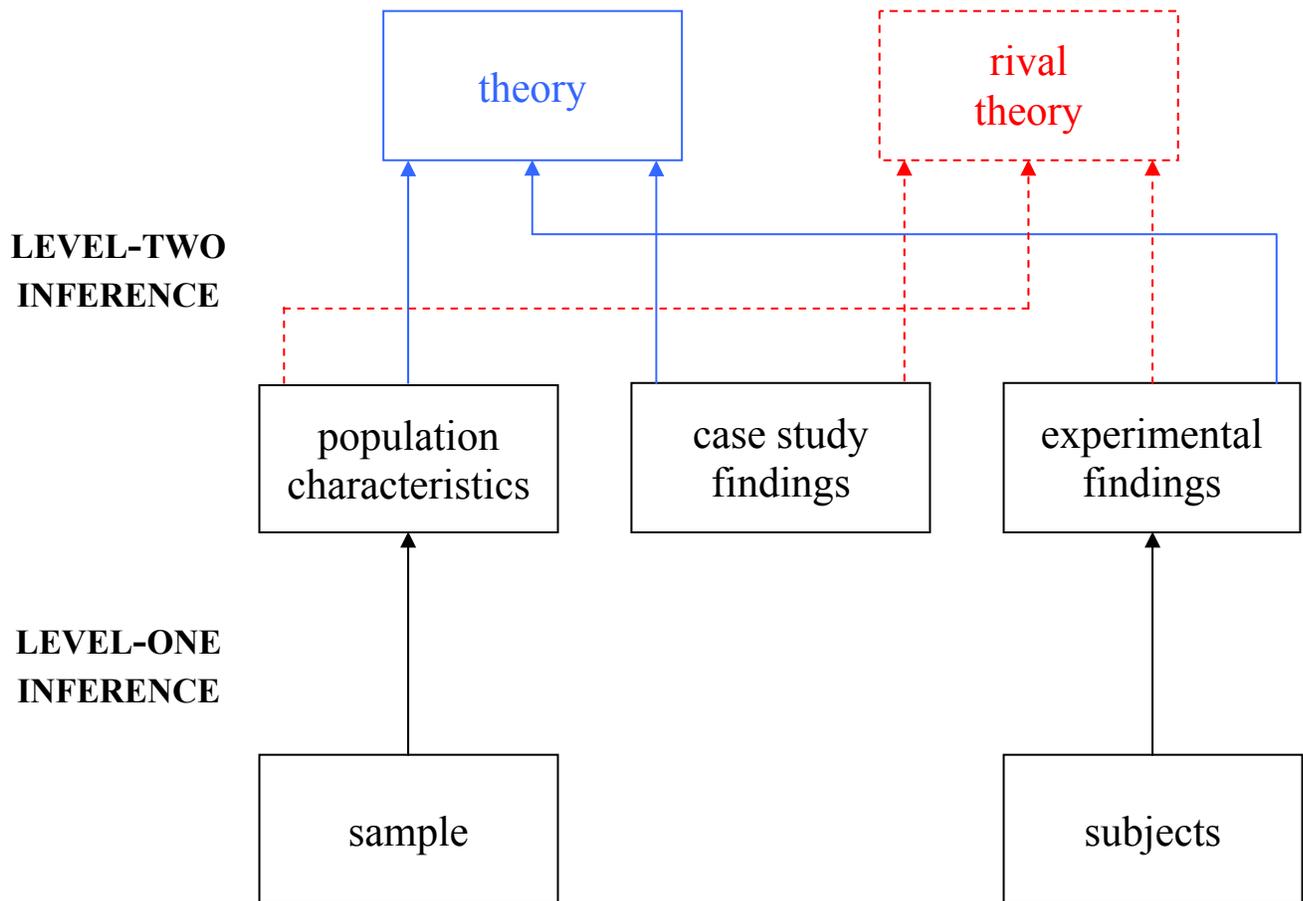
The examples in Tables 1 and 2 constitute evidence of the transference of a statistical, sampling-based conception of generalizability to research contexts that fall outside the boundaries of statistical sampling logic. In the next section of this paper, we review additional conceptions of generalizability.

III. Conceptions of Generalizability Beyond the Statistical

Methodologists familiar to IS researchers have already presented ideas about generalizability for non-statistical, non-sampling research situations. In this section of the paper, we examine four different forms of generalizability and four different forms of generalizing, based on concepts from Yin (1984, 1994), Campbell and Stanley (1963), Walsham (1995), Klein and Myers (1999), and Lee (1989a). The four forms of generalizing are: “generalizing from empirical statements,” “generalizing from theoretical statements,” “generalizing to empirical statements,” and “generalizing to theoretical statements.” The four forms of generalizability are “the generalizability of a theory to different settings,” “the generalizability of a theory within a setting,” “the generalizability of a variable, construct, or other concept to different theories,” and “the generalizability of a measure or observation to different settings.” The following discussion will provide the material that will appear in Figure 3, which categorizes and relates all of these different forms of generalizing and generalizability. A contribution of this paper is that no such overall conceptual framework or map about generalizability has been offered previously.

Yin

Yin’s conceptions about generalizing and generalizability are particularly important because they have influenced the conceptions that numerous other scholars have developed, including Walsham (1993) and Klein and Myers (1999), whom we discuss below. Furthermore, Yin’s monograph on case study research (first edition 1984, current edition 1994) is considered a “bible” among qualitative researchers. Yin (1994, pp. 31-32) draws a distinction between what he names level-one inference and level-two inference, where generalization takes different forms at the two levels (see Figure 2, below). In the following discussion, we add details and examples to Yin’s description.



(based on Figure 2.2 in Yin, 1994, p. 31)

Figure 2

For generalization in level-one inference, Yin gives two examples: generalizing from a sample to population characteristics (e.g., using the mean of a sample to measure the mean of the population from which the sample was taken) and generalizing from experimental subjects to experimental findings (e.g., using the test scores of research subjects to gauge the effect of an experimental treatment). We draw attention to a point that Yin does not mention: the product of generalization in level-one inference is empirical statements. In general, empirical statements refer to measurements or other observations of empirical or “real world” phenomena. Examples of empirical statements are numbers (such as a sample statistics) presented as measurements of characteristics of a population, a certain variable’s numerical value for an experimental effect

that a researcher predicts that she will observe after she administers the experimental treatment, and a case researcher's descriptions of the behaviors of organizational members.

As for generalization in level-two inference, Yin gives three examples, where all involve "generalizing to theory": generalizing from population characteristics to theory, generalizing from case study findings to theory, and generalizing from experimental findings to theory. Again, we draw attention to a point that Yin does not mention: here, unlike in level-one inference, the product of generalizing is not empirical statements; instead, generalizing begins with empirical statements and ends in theoretical statements. In general, theoretical statements posit the existence of phenomena and relationships that cannot be directly observed (and hence can only be theorized). Some beginning empirical statements could specify, for example, the measurements of the effect of a treatment administered in a particular field experiment; the rich details of a thick description 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, respectively, comprise 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' organizational culture that would help to account for the case study findings that were observed; or a theory explaining the underlying labor market forces that would result in the population characteristics that the sample estimated.

The terminology of "empirical statements" and "theoretical statements" is also useful for further detailing the concept of generalizability. In statistical generalization, the reasoning process proceeds from empirical statements (that specify the numerical values of the data points in a sample) to other empirical statements (that offer sample statistics as measurements of population characteristics). In contrast, when attempting to generalize a theory (e.g., a theory about a new technique for teaching reading that has been experimentally confirmed in one school district) to a new setting (e.g., a new school district), the inputs to the generalization process are theoretical statements (e.g., mathematical equations comprising the theory) and its outputs are empirical statements pertaining to a particular setting (e.g., specified values for the difference in students' reading abilities that the researchers predict that they will observe in the new school district if the theory is true). The process of generalizing from theoretical statements to empirical statements is routine among researchers. Hess and Kemerer (1994) generalize Malone *et al.*'s

“electronic markets hypothesis” (i.e., theoretical statements) to five different cases or field settings, resulting in particular descriptions (i.e., empirical statements) specifying what should and should not be observed in each of the five cases if the electronic markets hypothesis is true. Regarding the logic of generalizing a theory to a particular case setting, Markus (1989) offers guidelines for how to select a case for a disconfirmatory study. Yin’s own concept of “pattern matching,” which involves the matching of a theoretically derived pattern to the facts observed in a specific case, also involves the generalizing of an existing theory to a particular setting. Generalizing from theoretical statements to empirical statements is also acknowledged by Babbie (1990), who explains that when researchers seek this form of generalizability, they refer to the larger scope of empirical phenomena, beyond those directly captured in their immediate research.

Yin’s two levels of generalization are also useful for describing the larger research landscape in which statistical, sampling-based generalization is situated. In Yin’s conception of generalizability, statistical generalizing is certainly valid and useful, but the domain of its appropriate use is restricted: statistical generalizing is restricted to level-one inference and, even there, is just one way of how to generalize. In this conceptualization, statistical generalizing is but a subset or special case of generalizing. Equivalently stated, generalizing from a sample to a population does not constitute the general case of generalizing.

Yin gives the name “analytic generalization” to the different forms of generalizing that can occur in level-two inference. “In analytic generalization, the investigator is striving to generalize a particular set of results [such as case study findings or experimental findings] to some broader theory” (p. 36). “Analytic generalization” and “generalizing to theory” are synonyms.

Campbell and Stanley

Yin’s commentary on generalizability from a qualitative perspective complements Campbell and Stanley’s commentary on generalizability from a quantitative perspective. Campbell and Stanley’s classic book on experimental and quasi-experimental design (1963) uses the terms “external validity” and “generalizability” interchangeably (p. 5): “*External validity* asks the question of *generalizability*: To what populations, settings, treatment variables, and

measurement variables can this effect be generalized?” (p. 5). Campbell and Stanley’s research area is education. In their research area of education, we interpret the quotation as follows:

- “this effect” could refer to an improvement in students’ reading abilities observed in a field experiment that administers a new technique for teaching reading in a sample of students in a school district,
- a “population” to which this effect could be generalized would refer to the rest of the students in the school district from which the sample of students, making up the control and treatment groups, was taken,
- a “setting” to which this effect could be generalized could be a location, whether the remaining (i.e., the non-sampled) portion of the school district being studied or any school district outside the experiment,
- a “treatment variable” to which this effect could be generalized would refer to the new technique for teaching reading,
- and a “measurement variable” to which this effect could be generalized would refer to the the reading test by which reading abilities are measured for the treatment group’s students and the control group’s students, where measurements are taken both before and after the administering of the new technique for teaching reading to the experiment group’s students.

Hence, although Campbell and Stanley are well known for their emphasis on statistical inference in experimentation, their statement is also useful for emphasizing that they recognize not only generalizing to populations, but also generalizing to settings, generalizing to treatment variables, and generalizing to measurement variables.

Using their now familiar notation where “X” refers to a treatment and “O” refers to an observation or measure, Campbell and Stanley offer extended discussions (pp. 32-34) that they subtitle “Generalizing to Other Xs” and “Generalizing to Other Os.” Campbell and Stanley discuss instantiations of the same treatment X or measurement O within the same experiment. For instance, one concern would be whether the new technique for teaching reading is administered in the same way to all the students in the experiment’s treatment group; another concern would be whether the reading ability for each student, whether in the control group or treatment group, is measured in the same way. We recognize these concerns as pertaining to an experiment’s internal validity. As for the external validity of a treatment, X, we recognize that a different researcher could attempt to apply the same X (taking the form of an independent

variable) in a different theory; here, the generalizability or external validity of the treatment, X, would refer to the extent to which it can remain the same concept when transferred to the different theory. Correspondingly, as for the external validity of a measurement or observation, O, we recognize that a different researcher could attempt to make the same measurement or observation O (e.g., the measurement of the reading ability of students or the satisfaction of computer users) in a different setting; here, the generalizability or external validity of the measurement or observation, O, would refer to the extent to which it can be made in the same way in the different setting.

Another way to characterize Campbell and Stanley's conception of generalizability is to describe it as "transferability." For instance, will a study's conclusions remain valid if transferred to the rest of the population from which the study took its sample, to a real-world setting that its laboratory experiment only simulated, to an organization other than the one where the case researcher conducted his or her study, or to a new experiment that new researchers will be conducting in an attempt to replicate or extend the results of the original experiment? Will a treatment X still denote the same concept when transferred to a different theory? Will a measurement or observation O still be the made in the same way when the act of measuring or observing is transferred to a different researcher in a different study?

A point worth emphasizing is that the procedures of statistical inference, to which Campbell and Stanley give a prominent role in their conceptions of experimental testing, are valid for using a sample to make inferences about the same population from which the sample was taken, not a different population. Where "external validity" refers to the rest of the same population from which a sample is taken (i.e., the portion of the population external to the sample), increasing the size of the sample could improve sample-based inferences about the population. However, where "external validity" refers to a population different from the one from which the sample was taken, increasing the size of the sample would not improve inferences about the different population – a point to which we will return in the fourth section of this paper. There is no question that the procedures of statistical inference are valid, but they are valid only for making inferences about the population from which the sample was taken.

Table 3, below, draws on the distinction between empirical statements and theoretical statements in comparing Yin's conceptualization of generalization with Campbell and Stanley's.

We will return to this distinction in our later discussion where we critique statistical, sampling-based generalizability.

Table 3
Detailing and Comparing Yin’s and Campbell & Stanley’s
Conceptualizations of Generalization

	INPUT TO THE GENERALIZATION PROCESS	OUTPUT FROM THE GENERALIZATION PROCESS
GENERALIZING FROM A SAMPLE TO POPULATION CHARACTERISTICS ^{1,2}	empirical statements (e.g., the statements that specify the numerical values of the data points in a sample)	empirical statements (e.g., the statements that specify the sample statistics that are computed from the sample and used as measurements of population characteristics)
GENERALIZING FROM EXPERIMENTAL SUBJECTS TO EXPERIMENTAL FINDINGS ^{1,2}	empirical statements (e.g., the statements specifying the individual test scores of research subjects in the control group and treatment group, both before and after the administering of the experimental treatment)	empirical statements (e.g., the statement gauging the impact attributable to the treatment; what Campbell and Stanley call “the effect” of the treatment)
GENERALIZING A THEORY TO A DIFFERENT SETTING ^{1,3}	theoretical statements (e.g., the statements that comprise Malone <i>et al.</i> ’s electronic markets hypothesis)	empirical statements (e.g., the statements that Hess and Kemerer derive about what they should observe in each of five different cases if Malone <i>et al.</i> ’s electronic markets hypothesis is true)
GENERALIZING TO THEORY ⁴	empirical statements (e.g., the statements that specify the measurements of the effect of a treatment administered in a particular field experiment; the rich details of a thick description in a case study of a particular corporate headquarters; or the sample estimates of the population characteristics of workers in a particular geographic region)	theoretical statements (e.g., the statements that comprise a new theory positing new variables and the relationships among them that would account for the experimental effect that was measured in the field experiment; a theory explaining the corporate headquarters’ organizational culture that would help to account for the case study findings that were observed; or a theory explaining the underlying labor market forces that would result in the population characteristics that were estimated)

¹ consistent with the term “external validity” as used by Campbell & Stanley and by Yin

² mentioned by Yin as “level one inference”

³ consistent with Yin’s notion of “pattern matching”

⁴ mentioned by Yin as “level two inference” and also as “analytic generalization”

As Table 3 illustrates, Yin's "generalizing to theory" is a form of generalizing that is separate from and additional to Campbell and Stanley's "generalizing to populations" and "generalizing to settings." "Generalizing to populations" and "generalizing to settings" refer to generalizing to something *outside* of what the investigator is observing (e.g., generalizing to the portion of the population outside of the investigator's sample or to an altogether different population in a different setting), but "generalizing to theory" refers to generalizing *within* the particular empirical setting that the investigator is observing. As such, generalizing to theory happens to resonate with the anthropologist Geertz's view that "the essential task of [ethnographic] theory building is ... not to generalize across cases but to generalize within them" (1973, p. 26). In addition to Geertz, there are other notable scholars with whom Yin finds himself to be in good company regarding the concept of generalizing to theory. Glaser and Strauss (1967) and Strauss and Corbin (1998) have developed and advanced the research approach known as "grounded theory," which seeks the emergence of theory from data. The inductive case study approach of Eisenhardt (1989), which is built on the ideas of Yin as well as Glaser and Strauss, also involves generalizing to theory.

The notions of idiographic research and nomothetic research (Nagel, 1979, p. 547; Luthans and Davis, 1982) are pertinent here. The idea of generalizing within a setting is consistent with the idea of idiographic research, where the objective is to provide theoretical propositions that explain what idiographic researchers themselves would acknowledge to be unique events in unique circumstances. Fields such as history and anthropology frequently, if not mostly, engage in idiographic research. Idiographic research does not generalize outside, but within, the setting observed. Nomothetic research, in contrast, strives for theories that are true across a range of settings. For instance, physicists intend their theories to be valid not only on the earth, but in every part of the universe. Although generalizing to settings (nomothetic research) and generalizing to theory (idiographic research) are contrasting notions, they are not contradictory; they merely refer to different types of generalization, appropriate for different purposes and appropriate in different types of inquiry.

Walsham

Walsham, a noted proponent of interpretive case studies in IS research, refers to and builds on the work of Yin, particularly Yin's concept of "generalizing to theory." Walsham

(1995) specifies “four types of generalization from interpretive case studies” (p. 79). Explaining each type of generalizability by giving examples, Walsham illustrates (pp. 79-80) the “development of concepts” with Zuboff’s development of her concept of “informaté” from her case studies of information technology use in US organizations; the “generation of theory,” with Orlikowski and Robey’s drawing upon “their empirical work in IS to construct a theoretical framework concerned with the organizational consequences of information technology”; the “drawing of specific implications,” with Walsham and Waema’s drawing “a number of such implications based on an in-depth case study of the development of IS in a financial services company over an eight-year period”; and the “contribution of rich insight,” with Suchman’s case study of the use of a copying machine, where Suchman “provided rich insight on a range of topics, including the limits of machine intelligence, the inherent differences between plans and practical actions, and the need for more thoughtful machine design.” These four types of generalizability fall under the idea that “case studies ... are generalizable to theoretical propositions” (Yin, quoted in Walsham, p. 79).

Klein and Myers

Referring to and incorporating Walsham’s ideas, Klein and Myers’ “principle of abstraction and generalization” (1999) is one among a total of seven principles that they offer for how to conduct and evaluate interpretive field research in IS. Klein and Myers use the terms “generalization” and “theoretical abstraction” interchangeably. They state: “intrinsic to interpretive research is the attempt to relate particulars [observations of empirical details] ... to very abstract categories [theoretical statements]; unique instances can be related to ideas and concepts that apply to multiple situations.” As such, what Klein and Myers describe as the research act of abstracting or generalizing is consistent with Yin’s conception of analytic generalization or generalizing to theory. Interestingly, although Yin is usually classified as positivist and Walsham, Klein, and Myers, are considered to be interpretive, the concept of generalizing to theory is apparently accepted in both the positivist and interpretive approaches to research.

Lee

Lee (1989a) explains how the case study of an information system in an organization can satisfy the requirements of positivist science, even where the research involves qualitative observations of non-replicable events. As part of his discussion, Lee applies Campbell's (1975) three categories of "degrees of freedom," which Campbell takes beyond the statistical context traditionally associated with this term. In case research that tests a theory through its predictions, the categories pertain to, first, the number of different cases or organizational settings considered; second, the number of different predictions considered; and third, the number of rival theories considered.

An increase in each category or degree of freedom indicates an increase in a form of generalizability. First, consider the category of degrees of freedom pertaining to the number of different cases or organizational settings where a theory survives empirical testing. An increase in this number would establish the generalizability of the theory to these different settings. Note that this form of generalizability is independent of the size of any sample of observations within a single organizational setting; increasing such a sample size would increase the reliability or the statistical significance of sample-based measurements, but would not increase the generalizability of the tested theory to different settings. Where a theory-testing case study is conducted as a form of natural experiment (Lee, 1989b), this category of degrees of freedom refers to the generalizability that would be increased by increasing the number of different experiments, not the number of observations within a single experiment.

Second, consider the category of degrees of freedom pertaining to the number of different predictions that follow from a theory and that empirical testing confirms. (The different predictions of a theory, for instance, could refer to a new information system's ramifications for user acceptance, for corporate profitability, and for quality of working life.) In the study of a single case, an increase in the number of different predictions that follow from a theory and that empirical testing confirms would increase the generalizability of the theory within the setting being studied.

Third, regarding the category of degrees of freedom pertaining to the number of additional theories that are considered and that empirical testing in a single case setting would rule out, an increase in this number would likewise increase the generalizability of the surviving theory within the setting being studied.

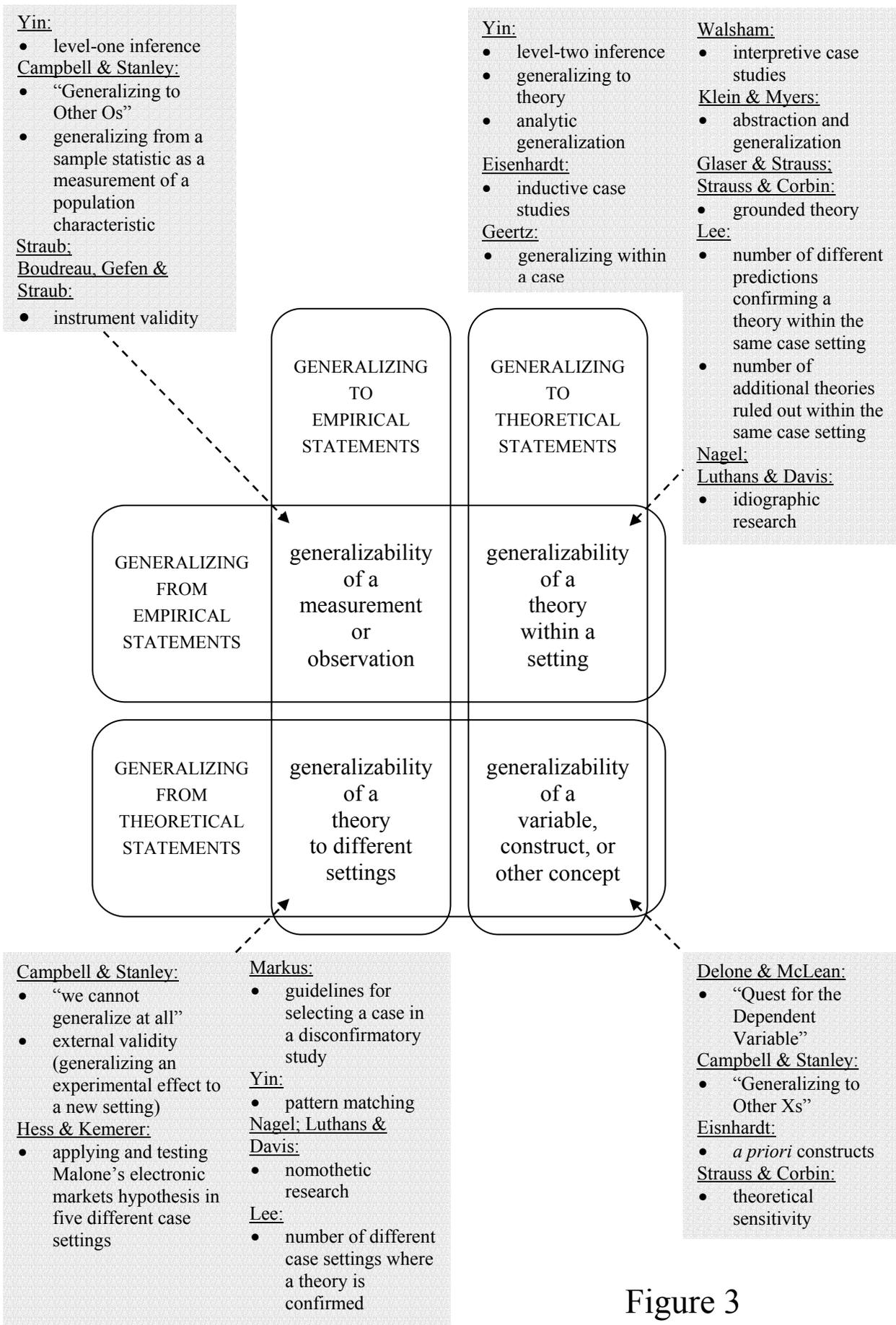


Figure 3

Recapitulation: Campbell and Stanley; Yin; Walsham; Klein and Myers; Lee

The preceding notions about generalizing and generalizability can be organized into the map shown in Figure 3 (above), which is an annotated version of Figure 1. The four ways in which Figures 1 and 3 use the term “generalizing” (“generalizing to empirical statements,” “generalizing to theoretical statements,” “generalizing from empirical statements,” and “generalizing from theoretical statements”) have already been discussed. The four ways in which it uses the term “generalizability” will now be elaborated. Explanations are given for annotations not already mentioned.

“Generalizability of a theory to different settings” is the extent to which a theory can remain valid when transferred to different settings, such as a new organization in contrast to the laboratory setting or field setting where the theory was studied. (We defer an explanation of the Campbell and Stanley quotation, “we cannot generalize at all,” to section IV.)

“Generalizability of a variable, construct, or other concept” is the extent to which a concept in a theory can remain the same when it is transferred to a different theory. In positivist research, an example of this would be a treatment variable, X, mentioned earlier regarding Campbell and Stanley’s “Generalizing to Other Xs.” Delone and McLean’s (1992) discussion of the quest for the dependent variable in IS research also falls in this category of generalizability. An interpretation of their argument is that if the same dependent variable (in particular, the success of an information system) cannot be transferred or generalized from the theory in one study to different theories in other studies, then the research findings across these studies cannot be considered cumulative. Another example of this category of generalizability would involve Eisenhardt’s acknowledgment that “*a priori* constructs” are needed and used in the generation of new theory; in other words, new theory is generalized, at least in part, using the constructs from other, existing theories. Strauss and Corbin’s similar acknowledgment of the importance of what they call “theoretical sensitivity,” which is the grounded theory researcher’s familiarity with the theory in the existing literature, likewise refers to the generalizing of new theory, at least in part, from existing theoretical concepts.

“Generalizability of a measurement or observation” is the extent to which a measurement or observation can be transferred to, or said to describe, a phenomenon other than the one that was actually measured or observed. Perhaps the most well known form of this is the generalizability of a measurement, obtained through a random sample, to the population from

which the sample was taken. Specifically, this refers to the extent to which the numerical value of a sample statistic (e.g., the sample mean or the estimated coefficient for an independent variable in a multivariate regression), computed from a random sample, can be considered a reliable measurement of the true value of the population characteristic for which it is but an estimate. If, in reality, the sample is not representative of the population (e.g., in Campbell and Stanley's education example, this could be a sample whose students all happen to come from only the highest socio-economic stratum in the population), then the resulting sample statistic would have a numerical value atypical for, and not generalizable to, the population. What statisticians call the confidence level or statistical significance is, in effect, a gauge of the extent to which a particular sample-based measurement can be considered typical for and generalizable to the population (in contrast to being considered atypical because of sampling error). As such, of the different forms of generalizability identified in Figure 3, the generalizability of a measurement or observation that is made through statistical random sampling is likely the best known and most developed form of generalizability.

The generalizability of a measurement or observation can be equivalently conceptualized as referring to the extent to which a measurement or observation (e.g., the measuring of the reading abilities of students or the satisfaction of computer users), made by one researcher in one setting, could be replicated by another researcher in a different setting. In positivist research, this conception of generalizability would characterize a measurement or observation that is made through a validated instrument (Straub, 1989; Boudreau, Gefen and Straub, 2001). Making a measurement or observation through a shared, validated instrument is one way of assuring that the same measurement or observation can be made by a different researcher. An instrument that has not been validated does not produce measurements that are replicable by or generalizable across different researchers making observations in different settings.

Finally, in Figure 3, "generalizability of a theory within a setting" refers to generalizability in idiographic studies, where the concern is about creating a theory applicable just to the setting being studied. This form of generalizability involves no statements about anything external to the setting examined in the study.

Note that a single study can involve different types of generalizability. For instance, a group support system (GSS) researcher conducting a laboratory experiment to test a theory can be concerned with the generalizability of her experimental treatment to other theories (which she

could promote by using a group facilitator who is trained in the same way that all group facilitators are trained for all GSS sessions, provided that such standards exist), the generalizability of her measurements or observations to other settings (which she could promote by using validated instruments), and the generalizability of her theory to other populations in other settings (which she could promote only if she is willing to make extra-logical assumptions, as we will discuss in section IV). In this GSS example, statistical sampling-based inference is useful and even critical in supplying the measurements needed in the separate and subsequent research step of testing the theory deductively through its predictions; however, we emphasize that these two forms of generalizability – (1) the generalizability of these measurements or observations, supplied through statistical sampling-based methods, to the rest of the population from which the sample was taken, and (2) the generalizability of the theory to settings outside of this particular researcher’s study – remain distinct.

In summary, Figure 3 shows, in a way complementary to Table 3, that statistical generalization (generalizing from a sample to its population) is but a subset of the general case of generalization.

Pertinence to Information Systems Research

IS research is well familiar with the generalizability of a measurement or observation (in the form of the generalizability of sample-based statistics to population characteristics) and has been concerned with the generalizability of a theory to different settings (in particular, the full variety of real-world business settings, in contrast to the setting of a single laboratory in an experiment or the setting of a single organization in a case study). However, the IS research studies in Tables 1 and 2, discussed earlier, do not distinguish between these two forms of generalizability. For instance, some of these studies cite a small sample size as the reason for a lack of generalizability to other organizations; this indicates a confounding of two different forms of generalizability because a small sample size pertains to the generalizability (or lack thereof) of a sample-based statistic to a population from which the sample is taken, rather than to the generalizability of a theory to different settings. (If we confine ourselves to the terminology of statistical sampling, then the generalizability of a theory to different settings would refer to the generalizability of a theory to different populations, not just to the population from which the sample was taken).

Furthermore, an increase in the number of organizations, sites, technologies, etc., that a given study examines can, but does not necessarily, increase the generalizability of the study's theory to different settings, as presumed by the instances of IS research in Tables 1 and 2. There are two reasons for this, as we will explain in detail in section IV.¹ It is noteworthy that two of the studies in Table 1 refer not only to the concept of generalizing to different settings, but also to the concept of generalizing to theory (even citing Yin). Significantly, however, their investigators diminish the value of this form of generalizability. Brown (1997) states "it should be reaffirmed that the single case research strategy employed here *only* allows generalizability to a research model" and Jarvenpaa and Leidner (1997) state "the single case site *limits* the generalizability of results" (italics are added in both quotations). Contrary to the limited or non-existent capability that Brown and that Jarvenpaa and Leidner believe their case research to have regarding generalizability, case research strategies not only support generalizing to theory, but also allow – subject to the same methodological conditions pertaining to laboratory experiments and statistical experiments (which we will examine in section IV) – generalizing to different settings.

Overall, the instances of IS research in Tables 1 and 2 presume statistical, sampling-based generalizing as the norm for generalizing and, hence, can be characterized as foregoing claims to greater generalizability prematurely and unnecessarily. We offer no negative criticism of the research conducted in any of the studies mentioned in Tables 1 and 2, except that the authors of these studies are being unnecessarily modest with respect to the claims of generalizability to which they are entitled.

Figure 3's map, which relates the different forms of generalizing and generalizability to one another, will be useful in our effort to formulate guidelines about how researchers may strive for and achieve generalizability in their research. However, before we offer these guidelines, we will offer a critique of statistical, sampling-based generalizability, so that we may better demarcate a proper place for it in our guidelines.

¹ We mention the two reasons briefly here. First, there is no logical basis for asserting that an increase in the number of observations in a study employing inductive generalizing provides stronger proof of the theory, much less a stronger claim to its generalizability to different settings. Second, even if an increase in the number of observations takes place in a framework that is deductive rather than inductive, there are still some additional qualifications that need to be satisfied in order for a researcher to claim an increase in the theory's generalizability to different settings.

IV. A Critique of Statistical, Sampling-Based Generalizability

Statistical generalizability – specifically, the generalizability of a sample to the population from which it is taken – is a valid concept. However, the uncritical application of statistical, sampling-based generalizability as the norm for all generalizability can lead to an improper assessment of a research study, where the improper assessment involves claiming (or denying) generalizability of a theory to different settings on the basis of sample size. Statistical generalization is a form of induction. After describing the difference between inductive logic and deductive logic, we will turn to existing critiques of induction to examine three known problems with induction, where any one of them is sufficient to rule out induction as a proper means of achieving the generalizability of a theory to different settings.

Inductive Logic and Deductive Logic

Inductive logic involves a reasoning process that begins with statements of particulars and ends in a general statement. The statements of particulars could refer to individual observations of a phenomenon and the general statement could be a general theory under which the individual observations would fall; here, the theory is said to be “induced” from the individual observations. In another example, the statements of particulars could refer to the numerical values of the data points in a random sample taken from a population, and the general statement offers the sample mean as a measurement of the population mean. Generalizing from a sample to a population is a form of induction. In inductive logic, generalizability is associated with the range of empirical circumstances covered by the statements of particulars from which the general statement is induced. For instance, increasing the size of a random sample would increase the generalizability of the resulting sample statistics as measurements of the corresponding population characteristics.

Deductive logic proceeds in the opposite direction: it involves a reasoning process that begins with a general statement, from which one or more statements of particulars are said to be “deduced.” The general statement could be a positivist theory of a phenomenon and the statements of particulars could be predictions, deduced from the theory, about how the phenomenon would behave when instantiated in the particular conditions of a specific setting. Along these lines, Lee (1989a) demonstrates how a case study by Markus (1983) performs the

testing of three different theories through their predictions. In a very different example, the general statement could refer to an interpretive theory (e.g., in the form of an ethnography) of an organization and the statements of particulars could refer to the interpretive researcher's anticipations (following from the ethnography) of the patterns of behaviors that particular organizational members would, or would not, instantiate in the particular conditions of a specific incident. In deductive generalizing, increasing the variety of different conditions in which a positivist theory is demonstrated to predict correctly (or an interpretive theory is demonstrated to lead to confirmed anticipations) could increase the generalizability of the theory (or the interpretation) to different settings.

First of Three Problems with Induction: the Infeasibility of Transforming Observation Statements into Theoretical Statements

An infeasibility associated with induction is that the induction of theory from data is not justified logically. It is a logical impossibility to generalize theoretical statements from empirical statements alone, or to transform empirical statements into theoretical statements. For example, examine Figure 3, where each row of data represents an observation or data point (such as a row of data that would be input to a statistical software such as SPSS, SAS, or even Excel):

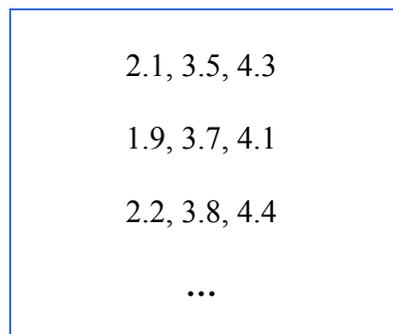


Figure 3

How would it be possible to “generalize” these data into a theory or to “induce” a theory from these data?

Consider the theory known as the Technology Acceptance Model. It posits three variables (perceived usefulness or U, perceived ease of use or EOU, and behavioral intention to use or BI) as well as the theoretical statements that “EOU has a direct effect on U,” “U has a

direct effect on BI,” and “EOU has a direct effect on BI.” There exists no mathematical or other logical procedure by which one would be able to “generalize” the numbers appearing in Figure 3 into any of the three variables (U, EOU, BI) or the three theoretical statements (“EOU has a direct effect on U,” “U has a direct effect on BI,” and “EOU has a direct effect on BI”). In other words, no mathematical procedure or other formal logic exists for transforming empirical statements into theoretical statements or, equivalently stated, for inducing or generalizing theoretical statements from empirical statements. For this reason, theory or “*a priori* constructs” are actually needed prior to data collection. Returning to our example, it is only after a researcher is aware of the theory (here, the Technology Acceptance Model) that the researcher can proceed to collect data (such as those appearing in Figure 3) to describe its instantiations.

In contrast, the induction of empirical statements from empirical statements – as in generalizing from a sample to population characteristics and generalizing from experimental subjects to experimental findings (see the discussion, above, accompanying Table 3) – is logically justifiable. Still, we emphasize that the generalizing of theory from data is logically impossible and that, therefore, increasing the size of a sample would not make the generalizing of theory from data any more possible.

The logical impossibility of transforming empirical statements into theoretical statements is not a new finding or novel argument. Kuhn states about himself and Karl Popper (1970a, p. 12): “...neither Sir Karl nor I is an inductivist. We do not believe that there are rules for inducing correct theories from facts, or even that theories, correct or incorrect, are induced at all.”

We note that social scientists, when formulating a theory, do indeed induce theory. In this paper, we have referred to it as “generalizing to theory” and “analytic generalization”; however, scientists do this by making guesses, not by using strict logic alone. Campbell and Stanley (p. 17) state: “*Logically*, we cannot generalize beyond these limits; i.e., we cannot generalize at all. But we do attempt generalization by guessing at laws...”

Second of Three Problems with Induction: the Same Evidence Can Support Contrary Theories

Another problem with induction is the fact that the same observation can be consistent with two different and incompatible theories. An observation that the sun rises in the morning and sets in the evening can support not only the Ptolemaic theory that the earth is the center of

the universe, but also the contrary Copernican theory that the earth revolves around the sun. Likewise, a data set that is consistent with Newtonian theory would also be consistent with Einstein's theory of relativity where, moreover, increasing the "sample size" of the data set would – if "inductive inference leads to valid theory" is momentarily accepted as a valid methodological principle – serve to strengthen the proof and generalizability of *both* theories, even though one of them is wrong. Because of this, increasing the size of a sample or the number of observations apparently consistent with a theory would not, by itself, make a theory any more correct or generalizable. Popper (1968a, pp. 37-38) provides illustrations in which supporters of astrology, supporters of the Marxist theory of history, and supporters of certain forms of psychoanalytic theories were able to provide more and more observations consistent with their theories, even though their theories were wrong.

Yin (1994) has also recognized this problem. In his Figure 2.2 (adapted as Figure 2 in this paper), he explicitly acknowledges the point that the same population characteristics can support inferences to both a theory and its rival theory, that the same case study findings can support inferences to both a theory and its rival theory, and that the same experimental findings can support inferences to both a theory and its rival theory. An increase in the number of observations supportive of both a theory and its rival theory would not, of course, better prove either theory or, hence, better establish either theory's generalizability.

Third of Three Problems with Induction: Hume's Truism

Campbell and Stanley (1963) themselves explicitly acknowledge that there exist "some painful problems in the science of induction" (p. 17) in their discussion of factors jeopardizing external validity. Because Campbell and Stanley are proponents of statistical analysis in experimentation, their cautionary words regarding induction take on added significance (*italics are in the original*):

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 4,² 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...

Because Campbell and Stanley refer to "Hume's truism," a closer look at Hume's truism is warranted.

Hume's discrediting of induction is a well accepted point in the philosophy and methodology of science. Rosenberg offers a succinct summary of this point (1993, p. 75):

Hume is almost universally credited with discovering the problem of induction...

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 are no more convincing than their most controversial premises and so generate a regress, while the inductive ones beg the question. Accordingly, claims that transcend available data, in particular predicts and general laws, remain unwarranted.

Along the same lines, Popper also points out that any attempt to justify inductive logic invokes an infinite regress in reasoning (1968b, p. 29):

² Campbell and Stanley describe "Design 4" as follows (1963, p. 13):

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

That inconsistencies may easily arise in connection with the principle of induction should have been clear from the work of Hume... For the principle of induction must be a universal statement in its turn. Thus if we try to regard its truth as known from experience, then the very same problems which occasioned its introduction will arise all over again. To justify it, we should have to employ inductive inferences; and to justify these we should have to assume an inductive principle of a higher order; and so on. Thus the attempt to base the principle of induction on experience breaks down, since it would lead to an infinite regress.

Hence, even apart from the first problem that the induction of theoretical statements from observation statements is logically impossible and the second problem that induction can support contrary theories, there is the third problem that induction itself is not justified logically. In this context, Campbell and Stanley's startling pronouncement, regarding the generalizability of a theory to different settings, is worth repeating: "we cannot generalize at all."

Pertinence to Information Systems Research

Any one of the three problems is sufficient to rule out induction – including the statistical, sampling-based conception of generalizability – as a justifiable means of achieving a theory's generalizability to different settings. The lesson from each problem is that, through logic alone (i.e., without the help of guesses, *a priori* constructs, and prior theories), the generalization of theory from observations is impossible. A larger number of observations (whether or not gathered through statistical random sampling) does not, in itself, result in an increase in the generalizability of a theory to different settings. As we will later discuss, an increase in the number of observations consistent with a theory can be useful, but is neither a necessary or sufficient condition for an increase in the theory's generalizability to different settings.

We emphasize that the three problems pertain to the generalization of theoretical statements from empirical statements, not the generalization of empirical statements from other empirical statements. There is no problem inherent in the latter. As such, the three problems do not invalidate the concept of statistical, sampling-based generalizability, which pertains to the generalizability of empirical statements (e.g., statements about sample statistics) to the

population being sampled; however, this concept does become invalid when it is improperly applied (or overgeneralized, as it were) to offer assessments of the generalizability of theoretical statements.

V. A Set of Guidelines for Achieving Generalizability

Fortunately, the philosophy and methodology of science has offered an alternative to induction. We will examine this alternative, from which we will draw guidelines for how to achieve what is arguably the form of generalizability that is the most problematic, but also the most important for information systems researchers: the generalizability of a theory to other settings. We will offer an illustration of how to apply the guidelines with material from Markus' classic 1983 study of power, politics, and MIS implementation. A reason for selecting Markus' study is that most readers are likely already familiar with it. This illustration of the guidelines will show how the logic of this paper can be applied in a proactive way to avoid confusion regarding generalizability. We will end this section with a brief mention of the methods, which already exist, for achieving the remaining three forms of generalizability.

Guidelines for Achieving the Generalizability of a Theory to Different Settings

Because there is no logical basis for inductively testing a theory or inductively establishing a theory's generalizability, the remaining alternative is for testing and generalizing to proceed deductively. Popper, in recognizing Hume's argument (discussed above), explains deductive justification in the context of his "criterion of demarcation" for distinguishing what may be properly considered science. According to this criterion, a theory must be stated in a form that would allow it to be proven wrong by evidence (i.e., "falsifiable," "disconfirmable," and "refutable"); the required logic for this, we will explain, is deductive, not inductive. The criterion of falsifiability is an enduring aspect of Karl Popper's philosophy – an aspect with which Thomas Kuhn registered his agreement even though he otherwise had major disagreements with Popper (1970b, p. 245): "...a field first gains maturity when provided with theory and technique which satisfy the four following conditions. First is Sir Karl's demarcation criterion without which no field is potentially a science..."

The logic of testing a theory by seeking out disconfirming observations is deductive in the following way. First, the theory is posited tentatively as true. Second, if true, then what the

theory predicts or hypothesizes to be observed in a particular laboratory experiment, field experiment, statistical experiment, etc., should then actually be observed. Hence this involves testing a theory through its logical deductions – which can be called predictions (if longitudinal) or hypotheses (if cross sectional). Under the conditions of controlled experimentation (or, in general, “controlled observation” [Nagel, 1979, p. 453]), actual observations that contradict the predictions or hypotheses deduced from the theory would then serve to disconfirm the theory.

According to this logic, a necessary condition for a theory to be considered scientific is not a large number of observations from which it could be induced or generalized, but its track record in surviving different attempts to marshal observations that would disconfirm it through rigorous experimentation involving deductive reasoning. For this reason, a scientific theory can never be properly described as “proven,” but can only be described, at best, as “corroborated” or “confirmed.” This means that the status of a theory as “valid” can be considered, at best, as only provisional. This also explains why some researchers prefer the locution that a theory is “consistent” with the evidence (the data, facts, or observations), rather than the locution that the evidence “proves” the theory. There always remains the possibility that another set of empirical circumstances, different from the particular empirical circumstances that the researcher observed in his or her experiment, sample, or case, can disconfirm the theory; the only way to find out would be to continue testing the theory, each time against a set of new and different empirical conditions.

Hence, in order to increase a theory’s generalizability in a deductive framework, a researcher would still need to make additional observations (which, for case research, could involve an increase in the number of cases or organizational settings where a theory is tested); however, this would have a meaning and function distinct from the act of increasing a sample size. First, statistical sampling is not deductive, but inductive; it involves observations in the form of empirical statements that are inputs or premises at the beginning of the reasoning process, where the outputs or conclusions would also be empirical statements (e.g., a statement offering a measurement of the impact of the experimental treatment). In contrast, in the deductive testing of a theory, observations play their critical role not at the beginning, but at the end of the reasoning process, where they are compared against deductions (predictions or hypotheses) that follow from theory. Second, statistical sampling is random, but the deductive

testing of a theory involves the targeted search for (not the random sampling of) those particular empirical conditions that would result in the most convincing disconfirmation of the theory.

Along these lines, Markus (1989, p. 24) explains that, in the deductive testing of a theory where the researcher is attempting to disconfirm a theory, it is actually proper and desirable for the researcher to seek out a setting (e.g., an organization or other field setting) where the theory is *most likely* to hold, in contrast to a setting that proponents of the theory would consider to be inhospitable or “stacked against” the theory. The reason is that, in the event that what the theory predicts (i.e., instantiations of particular actions, behaviors, and other phenomena) turns out to be contradicted by what the researcher observes, the disconfirmation of the theory would be more decisive. For instance, such a test of media richness theory could involve a setting that this theory itself would consider likely to support its prediction that “thin” media does not support rich communication – namely, a setting involving communication through e-mail; hence, observations revealing rich communication among e-mail users in a particular organization would provide a more convincing disconfirmation of media richness theory (see Markus, 1994; Lee, 1994; Ngwenyama and Lee, 1997). Denying the validity of a theory would also, of course, be denying its generalizability.

The logic of deductive testing also means that, strictly speaking, a theory cannot be generalized to settings where it has not yet been empirically tested and confirmed. The statement that “a theory, which has been confirmed through deductive testing in setting A, will also be valid in a new setting B” would itself be a theory requiring testing. We will use the term “strict generalizability of a theory to different settings” to refer to the restrictive position that the generalizability of a theory to different settings can be claimed only if the theory has been empirically tested, and confirmed, in each of the different settings to which it is claimed to be generalizable. Indeed, this position is so strict that, in effect, it has the same meaning as Campbell and Stanley’s conclusion that “we cannot generalize at all” – that is, we cannot generalize a theory at all to any settings where it has not yet been empirically tested or we can generalize a theory to a different setting only by actually testing and confirming it in that setting.

Less strictly speaking, a theory that has been confirmed through deductive testing in setting A can be generalized to a new setting B if one is willing to make three extra-logical assumptions. These less restrictive positions can be called “conservative generalizability” and “liberal generalizability,” depending on the extent to which the new setting (to which one wants

to generalize the theory) departs from the empirical conditions in the original setting where the theory was empirically confirmed. To help explain this, we continue with the earlier illustration involving Campbell and Stanley's domain of educational research. A claim of "conservative" or "liberal" generalizability would involve making the following three, related extra-logical assumptions:

- First, there is what Rosenberg calls the "uniformity of nature" assumption (quoted above) or what Campbell and Stanley call the "stickiness of nature" assumption: "the closer two events are in time, space, and measured value on any or all dimensions, the more they tend to follow the same laws" (Campbell and Stanley, p. 17). The generalization of a theory from one setting to another requires that this assumption be accepted.
- Second, there is what we will call the "successful identification of relevant variables" assumption. This is the assumption that all the same variables (e.g., age, intelligence, socioeconomic status, etc.) that were present in setting A and that were theorized as relevant to the shaping of the phenomena (e.g., the experimental effect, such as the difference in reading achievement scores attributable to the new experimental teaching technique) observed in setting A are also present and relevant in setting B. This also includes the assumption that any variables not identified (e.g., the magnetic field, barometric pressure, gamma radiation) are indeed irrelevant to the shaping of the phenomena that were observed in setting A and that will be observed in setting B.
- Third, there is the "sufficient similarity in relevant conditions" assumption. If a relevant condition is conceptualized as the particular value taken by a relevant variable, then this assumption would be that the value of this variable in setting B is sufficiently similar to its value in setting A, so as to support the assumption that setting B is "like" setting A. For instance, this could involve assuming that the students in setting B have the same levels of intelligence as the students in setting A (i.e., the distribution of IQ scores in setting B is sufficiently similar to the distribution of IQ scores in setting A) and come from the same socioeconomic backgrounds as the students in setting A (i.e., the distribution of household incomes in setting B is sufficiently similar to the distribution of household incomes in setting A).

A person willing to make all three of these extra-logical assumptions could then make the claim that a theory, empirically tested and confirmed in setting A, is generalizable to setting B, even

though it has not yet been tested and confirmed in setting B. Hence, researchers claiming either the “conservative” or “liberal” generalizability of a theory to a different setting would need to make the assumptions that they subscribe to the “uniformity” or “stickiness” of nature, that they have identified all the relevant variables, and that the relevant conditions (the values taken by the variables) in the new setting are sufficiently the same as in the setting where the theory was confirmed.

In addition to stating these assumptions, a researcher claiming the generalizability of his or her theory to a different setting should also clearly delineate, first, the values taken by the theory’s variables in the researched situation and, second, the difference in these values at the new setting(s) to which the theory is being generalized. A reader of the research could then take the responsibility for making his or her own decision as to whether the theory can be reasonably generalized to the new setting(s). This manner of reporting generalizability is no different from a statistical researcher’s reporting the p-value or statistical significance so that the reader can decide, for himself or herself, whether or not to consider the null hypothesis to be rejected (or, equivalently, reporting the researcher’s own arbitrarily chosen value of α as .10, .05, or .01 so that the reader can decide for himself or herself whether or not to agree with the researcher’s rejecting or not rejecting the null hypothesis). In both cases – the generalizing of a theory to a new setting and the testing of a hypothesis through statistical sampling – the reader or consumer of the research takes the responsibility for understanding the research and for making a judgment about whether to agree or disagree with the researcher. At the same time, it is the researcher’s responsibility to supply the information that the reader needs in order to make this decision properly.

Illustration

As a well known and well regarded empirical study, “Power, Politics, and MIS Implementation” (Markus, 1983) is a good choice for providing material with which to illustrate the guidelines for how to achieve the generalizability of a theory to different settings. “The web version of the Social Science Citation Index shows that over 200 other published studies have cited Markus’ classic since 1993 (the earliest year covered by the web version of the SSCI). Furthermore, its universal appeal is evident in its being regarded as an exemplar of not only

positivist research (Lee 1989a), but also interpretive research (Walsham 1993)” (Lee, Myers, Paré, Urquhart, and Markus, 2000). In this study, Markus tests three competing theories about user resistance to the implementation of management information systems (MIS) in a single field test, occurring in a real organization (which she calls “Golden Triangle Corporation” or GTC). The surviving theory is the “interaction theory,” the gist of which is that neither “people factors” alone (such as cognitive style, personality traits, and human nature) nor “system factors” alone (such as lack of user friendliness, lack of attention to human factors, and inadequate technical design) result in user resistance, but that interaction between the two sets of factors is a necessary condition for creating user resistance.

Based on Markus’ Figure 6 (p. 443) and her discussion accompanying it, we paraphrase Markus’ interaction theory as made up of the variables mentioned in Table 4a and the statements that follow Table 4a.

**Table 4a:
The Relevant Variables in
the Study where the Theory Was Empirically Confirmed**

Relevant variables (in Markus’ “Interaction Theory”)	OPD	“Organizational Power Distribution”
	ICB	“Intentions to Change Power Balance”
	PSSD	“Power Shift Implied in System Design”
	R	“Resistance” (by the intended users of a system to the system)
	PT	“Political Tactics” (for promoting and overcoming resistance)
	PSR	“Power Shift Realized in Organization”

The statements making up the Markus’ interaction theory (following from Markus’ Figure 6) are:

- OPD can have a direct influence on ICB.
- ICB can have a direct influence on PT.
- ICB can have a direct influence on PSSD.
- PT can have a direct influence on PSSD.
- PSSD can have a direct influence on R.
- R can have a direct influence on PSR.

- OPD can moderate the influence that R has on PSR.
- PT can moderate the influence that R has on PSR.

This theory (like any theory) consists of abstractions or general statements, not facts or actual empirical values describing any particular organizational setting. Markus applies the general statements of her theory to the Golden Triangle Corporation in the ways that Table 4b, below, indicates. Table 4b indicates the actual empirical values taken by the variables, identified in Table 4a, in the organizational setting of GTC.

**Table 4b:
The Relevant Conditions (the Variables and Their Values) in
the Study where the Theory was Empirically Confirmed**

Observed values of the variables at the GTC organization in Markus' field study	OPD	= GTC's divisional accountants have power to control access to financial data; GTC's corporate accountants lack power to control access to financial data.
	ICB	= The intention of Howard, head of the corporate controller's office, is to change the balance of power from the Chemical Divisions (where his rival, Spade, is Chemical Company Controller) to corporate management; corporate accountants intend "from the very beginning that FIS be used for managerial accounting, not just ... financial accounting" (p.439).
	PSSD	= The power shift in the system design of FIS is to give power to the corporate accountants and remove power from the divisional accountants.
	R	= The level of resistance from the divisional accountants to FIS was high.
	PT	= Corporate accountants had intentions of using FIS for managerial accounting but "did not immediately reveal these intentions to the divisions"; some divisional accountants continue using their old accounting methods ("thick manual ledger books") in parallel with FIS; some divisional accountants engage in "data fudging" when using FIS; corporate accounting makes adoption of FIS voluntary but requires non-adopters to report the same information manually.
	PSR	= Spade retires and his old position of Chemical Company Controller is "moved under the direct line control of corporate accounting and then eliminated," hence realizing a shift in power from divisional accounting to corporate accounting.

The facts which Markus observed at GTC (and which are displayed in Table 4b) can be said to "populate" the variables and statements of her interaction theory, much as the

instantiation of a database at a point in time can be said to populate its database schema.³ Now, to what extent can Markus' interaction theory be generalized to organizational settings other than GTC?

According to the position of strict generalizability, one may not say at all that Markus' interaction theory is valid for any organizational setting other than GTC, unless one actually applies, empirically tests, and confirms it in the other organizational setting.

According to the positions of conservative generalizability and liberal generalizability, one may properly say that Markus' interaction theory is valid for an organizational setting other than GTC only if one accepts the three assumptions (described above, which are the "uniformity of nature" assumption, the "successful identification of relevant variables" assumption, and the "sufficient similarity in relevant conditions" assumption). One makes the claim of conservative generalizability if the values of the relevant variables in the new organizational setting, to which one wishes to generalize the theory, are "close to" or "only slightly different from" their values in the original setting where it was empirically tested and confirmed. One makes the claim of liberal generalizability if the values of the relevant variables in the new organizational setting, to which one wishes to generalize the theory, are "far from" or "much different from" their values in the original setting where it was empirically tested and confirmed. Tables 5 and 6 illustrate conservative and liberal generalizability. Because a researcher would make explicit, in his or her report, the extent to which the values of the variables in the new setting diverge from the values that they had in the original setting where the theory was tested and confirmed, a reader could then decide for himself or herself whether or not to accept the generalizability (whether conservative or liberal) that the research is imputing to the theory.

Table 5 uses the color red to indicate a variable's change from its original value (designated with a "strikethrough font") at the GTC organizational setting (where Markus empirically tested and confirmed the interaction theory) to the new value at the different organizational setting to which the interaction theory is being generalized.

³ A corresponding table could be constructed to illustrate the application of the interaction theory's statements (see the bullet points above) to GTC. For instance, the application of the theoretical statement "OPD can have a direct influence on ICB" would be "GTC's organizational power distribution, which favors the divisional accounts over the corporate accountants, can foster or increase the intentions of Howard, head of the corporate controller's office, to change the balance of power away from the divisions." Similarly, the application of the theoretical statement "PSSD can have a direct influence on R" would be "The power shift implied in the system design of FIS can engender user resistance."

**Table 5:
Conservative Generalizability**

A set of values allowed* for the relevant variables in other organizations to which Markus' interaction theory is generalizable (under the assumption of conservative generalizability)	OPD	= GTC's divisional accountants managers in one corporate division have power to control access to financial valued data (henceforth, this corporate division with such power is abbreviated as "CD-power"); GTC's corporate accountants managers in another corporate division lack power to control access to financial valued data ("CD-no-power").
	ICB	= The intention of Howard, head of the corporate controller's office , a key player in top management is to change the balance of power from the Chemical Divisions CD-power (where his or her rival, Spade, is Chemical Company Controller is head) to corporate management the CD-no-power; corporate accountants managers in the CD-no-power intend "from the very beginning that FIS the new accounting system be used for managerial accounting accessing the valued data directly, not just ... financial accounting."
	PSSD	= The power shift in the system design of FIS the new accounting system is to give power to the corporate accountants managers in the CD-no-power and remove power from the managers in the CD-power .
	PT	= Corporate accountants managers in the CD-no-power had intentions of using FIS the new accounting system for managerial accounting accessing the valued data directly but "did not immediately reveal these intentions to the divisions CD-power"; some divisional accountants managers in the CD-power continue using their old accounting methods ("thick manual ledger books") in parallel with FIS the new accounting system; some divisional accountants managers in the CD-power engage in "data fudging" when using FIS the new accounting system; corporate accounting the CD-no-power makes adoption of FIS the new accounting system voluntary but requires non-adopters to report the same information manually.
	R	= The level of resistance from the divisional accountants managers in the CD-power to FIS the new accounting system was is high.
	PSR	= Spade the head of the CD-power retires and his or her old position of Chemical Company Controller is "moved under the direct line control of corporate accounting the CD-no-power and then eliminated," hence realizing a shift in power from divisional accounting the CD-power to corporate accounting the CD-no-power.

*These values depend on how "literal," "conservative," or "liberal" one desires the generalizability to be.

Thus a reader can see that the (conservative) generalization is from "GTC's divisional accountants" to, in general, "managers in one corporate division"; from "financial data" to "valued data" in general; from "Howard, head of the corporate controller's office" to any "key player in top management"; from "Chemical Divisions" to "corporate divisions with power";

from the specific function of “managerial accounting” to the more general function of “accessing the valued data directly”; and from the “system design of FIS” to the “system design of the new accounting system” in general. The reader, especially if a manager in a company, could then make the decision for himself or herself (1) whether or not the conservative generalization diverges too much from the original circumstances where the theory was empirically confirmed and (2) whether or not the conservative generalization captures the circumstances at his or her own organization.

Table 6, below, uses a “double strikethrough” font to indicate the change from the earlier, conservatively generalized value of a variable to a more liberally generalized value – one that departs even further from the value observed in the setting where the researcher empirically confirmed the theory. A reader could easily see, by following the progression from the single strikethrough font to the double strikethrough font to the plain red font, the extent of the change from the original setting.

**Table 6:
Liberal Generalizability**

A set of values allowed* for the relevant variables in other organizations to which Markus’ interaction theory is generalizable (under the assumption of liberal general-	OPD	= GTC’s divisional accountants managers in one corporate division members of one stakeholder group have power to control access to financial valued data (henceforth, this stakeholder group is abbreviated as “SG-power”); GTC’s corporate accountants managers in another corporate division members of another stakeholder group lack power to control access to financial valued data (“SG-no-power”).
	ICB	= The intention of Howard, head of the corporate controller’s office, a key player in top management a key player who is a member of SG-no-power is to change the balance of power from the Chemical Divisions CD power SG-power (where his or her rival, Spade, is Chemical Company Controller is head) to corporate management the CD no-power SG-no-power; corporate accountants managers in the CD no-power members of the SG-no-power intend “from the very beginning that FIS the new accounting information system be used for managerial accounting accessing the valued data directly shifting the organizational balance of power to themselves, not just ... financial accounting accessing data. ”
	PSSD	= The power shift in the system design of FIS the new accounting information system is to give power to the corporate accountants managers in the CD no-power members of the SG-no-power and remove power from the managers in the CD power members of the SG-power.

izability)	PT	= Corporate accountants managers in the CD no power members of the SG-no-power had intentions of using FIS the new accounting information system for managerial accounting accessing the valued data directly shifting the organizational balance of power to themselves but “did not immediately reveal these intentions to the divisions CD power SG-power”; some divisional accountants managers in the CD power members of the SG-power continue using their old accounting methods (“ thick manual ledger books ”) in parallel with FIS new accounting information system; some divisional accountants managers in the CD power members of the SG-power engage in “data fudging” when using FIS the new accounting information system; corporate accounting the CD no power SG-no-power makes adoption of FIS new accounting information system voluntary but requires non-adopters to report the same information manually.
	R	= The level of resistance from the divisional accountants managers in the CD power members of the SG-power to FIS the new accounting information system was is high.
	PSR	= Spade the head of the CD power SG-power retires and his or her old position of Chemical Company Controller is “moved under the direct line control of corporate accounting the CD no power SG-no-power and then eliminated,” hence realizing a shift in power from divisional accounting the CD power SG-power to corporate accounting the CD no power SG-no-power.

*These values depend on how “literal,” “conservative,” or “liberal” one desires the generalizability to be.

For instance, “managers in one corporate division” becomes generalized to “members of one stakeholder group” while “ a key player in top management” is generalized to “a key player who is a member of a stakeholder group that lacks power to control access to valued data.” As in the case of conservative generalizability, the researcher’s generalization is sufficiently explicit so that a reader can take responsibility for accepting the generalization and applying it to his or her own organization. Again, this is no different from a statistical researcher’s explicitly stating the p-value associated with the test of a particular hypothesis, so that a reader can decide for himself or herself whether the hypothesis should be rejected.

Table 7, below, offers an alternative and more succinct way of conveying the same information in Tables 5 and 6.

**Table 7:
Summary of the Conservative and Liberal Generalizing
of Markus' Theory to Different Settings**

THE RELEVANT VARIABLES	THE VALUES OF THE RELEVANT VARIABLES IN THE SETTING WHERE THE THEORY WAS EMPIRICALLY CONFIRMED	THE VALUES OF THE RELEVANT VARIABLES IN A DIFFERENT SETTING TO WHICH THE THEORY IS BEING CONSERVATIVELY GENERALIZED	THE VALUES OF THE RELEVANT VARIABLES IN A DIFFERENT SETTING TO WHICH THE THEORY IS BEING LIBERALLY GENERALIZED
OPD =	...GTC's divisional accountants have power to control access to financial data; GTC's corporate accountants lack power to control access to financial data.	...managers in one corporate division have power to control access to valued data (henceforth, this corporate division with such power is abbreviated as "CD-power"); managers in another corporate division lack power to control access to valued data ("CD-no-power").	...members of one stakeholder group have power to control access to valued data (henceforth, this stakeholder group is abbreviated as "SG-power"); members of another stakeholder group lack power to control access to valued data ("SG-no-power").
ICB =	...The intention of Howard, head of the corporate controller's office, is to change the balance of power from the Chemical Divisions (where his rival, Spade, is Chemical Company Controller) to corporate management; corporate accountants intend "from the very beginning that FIS be used for managerial accounting, not just ... financial accounting" (p.439).	...The intention of a key player in top management is to change the balance of power from the CD-power (where his or her rival is head) to the CD-no-power; managers in the CD-no-power intend "from the very beginning that the new accounting system be used for accessing the valued data directly, not just ... accounting."	...The intention of a key player who is a member of the SG-no-power is to change the balance of power from the SG-power (where his or her rival is head) to the SG-no-power; members of the SG-no-power intend "from the very beginning that the new information system be used for shifting the organizational balance of power to themselves, not just ... accessing data."

PSSD =	...The power shift in the system design of FIS is to give power to the corporate accountants and remove power from the divisional accountants.	...The power shift in the system design of the new accounting system is to give power to the managers in the CD-no-power and remove power from the managers in the CD-power.	...The power shift in the system design of the new information system is to give power to the members of the SG-no-power and remove power from the members of the SG-power.
R =	...The level of resistance from the divisional accountants to FIS was high.	...The level of resistance from the managers in the CD-power to the new accounting system is high.	...The level of resistance from the members of the SG-power to the new information system is high.
PT =	...Corporate accountants had intentions of using FIS for managerial accounting but "did not immediately reveal these intentions to the divisions"; some divisional accountants continue using their old accounting methods ("thick manual ledger books") in parallel with FIS; some divisional accountants engage in "data fudging" when using FIS; corporate accounting makes adoption of FIS voluntary but requires non-adopters to report the same information manually.	...managers in the CD-no-power had intentions of using the new accounting system for accessing the valued data directly but "did not immediately reveal these intentions to the CD-power"; some managers in the CD-power continue using their old accounting methods in parallel with the new accounting system; some managers in the CD-power engage in "data fudging" when using the new accounting system; the CD-no-power makes adoption of the new accounting system voluntary but requires non-adopters to report the same information manually.	...members of the SG-no-power had intentions of using the new information system for shifting the organizational balance of power to themselves but "did not immediately reveal these intentions to the SG-power"; some members of the SG-power continue using their old methods in parallel with new information system; some members of the SG-power engage in "data fudging" when using the new information system; the SG-no-power makes adoption of new information system voluntary but requires non-adopters to report the same information manually.
PSR =	...Spade retires and his old position of Chemical Company Controller is	...the head of the CD-power retires and his or her old position is "moved under the direct	...the head of the SG-power retires and his or her old position is "moved under the direct

	<p>“moved under the direct line control of corporate accounting and then eliminated,” hence realizing a shift in power from divisional accounting to corporate accounting.</p>	<p>line control of the CD-no-power and then eliminated,” hence realizing a shift in power from the CD-power to the CD-no-power.</p>	<p>line control of SG-no-power and then eliminated,” hence realizing a shift in power from the SG-power to the SG-no-power.</p>
--	--	---	---

In general, the research act of showing a theory’s generalizability to different settings can involve, in addition to making the three assumptions mentioned above, a careful identification and explicit presentation of the theory’s variables, the values taken by the variables in the setting where they theory was empirically confirmed, and the different values that the variables would take in the new setting to which the theory is being generalized.

Guidelines for Achieving the Remaining Three Forms of Generalizability

Guidelines for achieving each of the three remaining forms of generalizability are better known and require little additional elaboration.

The generalizability of a measurement or observation. The procedures of statistical inference (classic texts include Neter, Wasserman, and Whitmore, 1993, and Wonnacott and Wonnacott, 1990) explain how to generalize sample-based measurements (e.g., a sample mean or a sample standard deviation) to their corresponding population characteristics (e.g., the population mean or the population standard deviation). As for the procedures of instrument validation, Smith, Milberg, and Burke (1996, p. 188) explain that an instrument “must be validated with different populations” so that measurements subsequently made through it can be said to be valid across different subjects, groups, conditions, settings, and times (also see Straub, 1989, and Boudreau, Gefen, and Straub, 2001). The procedures of statistical inference and the procedures of instrument validation are well developed, well known, and well accepted.

The generalizability of a variable, construct, or other concept. The generalizability issue here pertains to a research community’s consistent application of its research practices. DeLone and McLean (1992) examine one dimension of this. They observe six different ways in which different IS research studies have conceptualized the dependent variable of IS success;

they are system quality, information quality, use, user satisfaction, individual impact, and organizational impact (where, in another layer of complication, each of these six ways of conceptualizing IS success has been measured in different ways in different studies). This impedes the generalizability of the concept of IS success across studies. Campbell and Stanley (1963) examine another dimension of this. Regarding the generalizing of an independent or treatment variable from one research study to another, Campbell and Stanley point out (in their discussion, under the heading “Generalizing to Other Xs,” pp. 32-33) that generalizability follows from the extent to which researchers execute a given treatment (e.g., the administering of a new technique for teaching reading) in the same way from one study to the next. In both the DeLone and McLean case and the Campbell and Stanley case, the generalizability of a variable, construct, or other concept pertains not so much to any logic as it does to how consistently a research community applies its research practices.

The generalizability of a theory within a setting. Interpretive research, especially ethnography, has established procedures for validating and generalizing a theory with a given setting. These procedures are well known both inside and outside of IS research (Gearing, 1970; Geertz, 1973; Kanter, 1977; Agar, 1986; Golden-Biddle and Locke, 1993; Walsham and Sahay, 1999; Klein and Myers, 1999; Schultze, 2000).

VI. Discussion

In qualitative and case research, the generalizing of a theory to different settings rests on assumptions. As such, this form of generalizing finds itself in good company: the generalizing of quantitative and statistical results similarly rests on assumptions. We emphasize that, throughout this paper, we have recognized the legitimacy of statistical generalizability – specifically, generalizing from a statistical, sample-based measurement (such as the sample mean, or a regression coefficient calibrated from the sample data) to a characteristic of the population (such as the population mean or the true value of the regression coefficient) from which the sample was taken. The assumptions on which this well known and widely practiced form of generalizing rests are useful for providing a parallel context with which to elucidate the generalizing of a theory to different settings.

Consider the meaning of a statistical test of a null hypothesis where the sample-based observations result in the rejection of the null hypothesis and where the level of statistical

statistical significance = α
 = “ p-value ”
 = $P(\text{ type I error } | H_0)$

Increasing the sample size will improve the numerical value of α , the p-value, and $P(\text{ type I error } | H_0)$, *but* increasing the sample size will *not* provide any additional measure of proof of H_0 or make H_0 any more generalizable.

an *assumption*:
 “on the condition that”
 or “assuming that”
 or “given that” H_0 is true.

Figure 4

significance associated with this particular sample is reported to be .01. This is conventionally considered to be an excellent level of statistical significance, but this does not mean that the null hypothesis is true or generalizable. Strictly speaking, this number is a conditional probability; it means that *if, on the condition that, assuming that, or given that* the null hypothesis is true, then the probability that the sample-based observations would be sufficiently skewed so as to result in (incorrectly) rejecting the null hypothesis is only one percent. In this example, the excellent level of statistical significance does not establish, but actually assumes, that the null hypothesis is true. Increasing the sample size would improve the reported statistical significance and its reliability (the extent to which other random samples of the same size would lead to the same conclusion of rejecting or not rejecting the null hypothesis), but would not provide any additional measure of proof of the null hypothesis or make it any more generalizable; in other words, the truth of the null hypothesis (and hence its generalizability to different settings) would remain assumed and undemonstrated, as summarized in Figure 4.

With regard to generalizing a theory to different settings, Tables 5 and 6 similarly rest on the assumption that the theory being tested is true (to the extent that it has not yet been disproved) and will remain true (i.e., not disproved) even when the values of the variables change to the extent that Tables 5 and 6 explicitly indicate. And just as the threshold for an acceptable level of statistical significance (e.g., .10, .05, .01, etc.) is a matter for convention and judgment, the threshold for an acceptable change in the value of a variable from the situation where the theory was empirically confirmed (e.g., how conservative must, or how liberal may, this change in value be) is likewise a matter for convention and judgment.

This line of reasoning leads to the following question: if quantitative and statistical research may not claim an increase in the generalizability of a theory to different settings on the basis of an increase in sample size, then how may quantitative and statistical research be allowed

to claim the generalizability of a theory to different settings at all? Note that all of the arguments presented in this paper pertain just as well to quantitatively and statistically performed research as to qualitative and case research. Whether a theory is stated mathematically (for example, in the form of equations made up of independent, dependent, and moderating variables) or qualitatively (in the form of words), the same arguments about generalizing from empirical or theoretical statements and generalizing to empirical or theoretical statements still apply. Tables 4a, 4b, 5, 6, and 7, which operationalize the generalizing of Markus' interaction theory to settings other than Golden Triangle Corporation, provide an illustration of this point; these tables' actions of explicitly identifying, defining, and populating a theory's variables apply equally well to qualitatively stated theories (such as Markus') and to mathematically stated theories. Furthermore, the three required antecedent assumptions (the "uniformity of nature" assumption, the "successful identification of relevant variables" assumption, and the "sufficient similarity in relevant conditions" assumption) similarly make no discrimination between quantitative and qualitative research.

Our reasoning in this paper has, in effect, generalized the conception of how to achieve the generalizability of a theory to different settings so that it pertains as much to quantitative and statistical research as it does to qualitative and case research.

VII. Conclusion

The concept of generalizability itself is not homogeneous and monolithic. It refers to four different forms of generalizability (the generalizability of a theory to different settings, the generalizability of a theory within a setting, the generalizability of a measurement or observation, and the measurement of a variable, construct, or other concept), where each one of them results from a different combination of different types of generalizing (generalizing from or to empirical statements and generalizing from or to theoretical statements). There is no justification for the current and widespread IS research practice of imposing the statistical, sampling-based conception of generalizability (which falls under the generalizability of a measurement or observation) as the norm for either judging or prescribing qualitative research efforts to achieve the generalizability of a theory to different settings. No one of the four forms of generalizability has an inherent primacy over any other. Each is legitimate for its own purposes.

The overgeneralizing of the statistical, sampling-based conception of generalizability to qualitative and case research that seeks to generalize a theory to different settings is not simply an academic error of no consequence; it is harmful. It improperly thwarts the development and publishing of good theory, hence resulting in the research audience's being denied the opportunity to reach an informed understanding of the theory and to take responsibility for making their own judgment about the theory's generalizability to their own settings. The discipline of IS research, unlike the pure sciences, also has professional constituents (IS executives, managers, and consultants) and operates in concert with technologies that are relentlessly advancing. Our understanding of the social, organizational, and managerial aspects of our field may be trailing far behind our understanding of the technological issues. Unnecessarily confining the application of new theories from qualitative and case research to new settings can cripple our ability to keep pace. Applying a variety of forms of generalizability promises to improve the development of more current and more useful social, organizational, and managerial theory.

References

- Agar, M. (1986) *Speaking of Ethnography*, Sage Publications, Beverly Hills, CA.
- Albers, M., Agarwal, R., and Tanniru, M. (1994) "The Practice of Business Process Reengineering: Radical Planning and Incremental Implementation in an IS Organization" in *SIGCPR '94. Proceedings of the 1994 Computer Personnel Research Conference on Reinventing IS: Managing Information Technology in Changing Organizations*, ACM Press, Arlington, VA, 87-96.
- Babbie, E. (1990) *Survey Research Methods 2nd Ed.*, Wadsworth, Belmont, California.
- Boudreau, M., Gefen, D., and Straub, D. (2001) "Validation in Information Systems Research: A State-of-the-Art Assessment." *MIS Quarterly*, 25 (1), 1-16.
- Brown, C. (1997) "Examining the Emergence of Hybrid IS Governance Solutions: Evidence from a Single Case Site." *Information Systems Research*, 8 (1), 69-94.
- Campbell, D. (1975) "'Degrees of Freedom' and the Case Study." *Comparative Political Studies*, 8 (2), 178-193.

- Campbell, D. and Stanley, J. (1963) *Experimental and Quasi-Experimental Designs for Research*, Houghton Mifflin, Boston, MA.
- DeLone, W. and McLean, E. (1992) "Information Systems Success: the Quest for the Dependent Variable." *Information Systems Research*, 3 (1), 60-95.
- Eisenhardt, K. (1989) "Building Theories from Case Study Research.." *Academy of Management Review*, 14 (4), 532-550.
- Gearing, F. (1970) *The Face of the Fox*, Aldine Publishing, Chicago, IL.
- Geertz, C. (1983) "'From the Native's Point of View': On the Nature of Anthropological Understanding" in *Local Knowledge* (Geertz, C., ed.), Basic Books, New York, 55-70.
- Gefen, D. and Straub, D. (1997) "Gender Differences in the Perception and Use of E-mail: An Extension to the Technology Acceptance Model." *MIS Quarterly*, 21 (4), 389-400.
- Glaser, B. and Strauss, A. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine Publishing, Chicago, IL.
- Golden-Biddle, K. and Locke, K. (1993) "Appealing Work: An Investigation of How Ethnographic Texts Convince," *Organization Science*, 4 (4), 595-616.
- Goodhue, D. and Thompson, R. (1995) "Task-Technology Fit and Individual Performance." *MIS Quarterly*, 19 (2), 213-226.
- Hess, C. and Kemerer, C. (1994) "Computerized Loan Origination Systems: An Industry Case Study of the Electronic Markets Hypothesis." *MIS Quarterly*, 18 (3), 251-275.
- Hidding, G. (1998) "Adoption of IS Development Methods across Cultural Boundaries" in *Proceedings of the Nineteenth International Conference on Information Systems* (Hirschheim, R., Newman, M., and DeGross, J., eds.), 308-312.
- Jarvenpaa, S. L., and Leidner, D. (1997) "An Information Company in Mexico: Extending the Resource-Based View of the Firm" in *Proceedings of the Eighteenth International Conference on Information Systems*, (Kumar, K. and DeGross, J., eds.), 75-87.
- Klein, H. and Myers, M. (1999) "A Set of Principles for Conducting and Evaluating Field Studies in Information Systems." *MIS Quarterly*, 23 (1), 67-93.
- Kuhn, T. (1970a) "Logic of Discovery or Psychology of Research?" in *Criticism and the Growth of Knowledge* (Lakatos, I. and Musgrave, A., eds.), Cambridge University Press, New York, 1-23.

- Kuhn, T. (1970b) "Reflections on My Critics" in *Criticism and the Growth of Knowledge* (Lakatos, I. and Musgrave, A., eds.), Cambridge University Press, New York, 231-278.
- Lee, A. (1989a) "A Scientific Methodology for MIS Case Studies." *MIS Quarterly*, 13 (1), 33-50.
- Lee, A. (1989b) "Case Studies as Natural Experiments." *Human Relations*, 42 (2), 117-137.
- Lee, A. (1994) "Electronic Mail as a Medium for Rich Communication: An Empirical Investigation Using Hermeneutic Interpretation." *MIS Quarterly*, 18 (2), 143-157.
- Lee, A. (1999) "Researching MIS," in *Rethinking MIS* (Currie, W. and Galliers, R., eds.), Oxford University Press, New York, 7-27.
- Lee, A., Myers, M., Paré, G., Urquhart, C., Markus, M.L. (2000) "Three Perspectives: If Markus' 1983 Classic Study, 'Power, Politics, and MIS Implementation,' Were Being Reviewed Today," panel at the Twenty-First International Conference on Information Systems, selected panel materials available at <http://www.people.vcu.edu/~aslee/ICIS2000/ICIS2000-Panel-Markus1983.htm>.
- Luthans, F. and Davis, T. (1982) "An Idiographic Approach to Organizational Behavior Research: The Use of Single Case Experimental Designs and Direct Measures." *Academy of Management Review*, 7 (3), 380-391.
- Majchrzak, A., Rice, R., Malhotra, A., King, N., and Sulin, B. (2000) "Technology Adaptation: The Case of a Computer-Supported Inter-organizational Virtual Team." *MIS Quarterly*, 24 (4), 569-600.
- Malone, T., Yates, J., and Benjamin, R. (1987) "Electronic Markets and Electronic Hierarchies." *Communications of the ACM*, 30 (6), 484-497.
- Markus, M.L. (1983) "Power, Politics, and MIS Implementation." *Communications of the ACM*, 26 (6), 430-444.
- Markus, M.L. (1989) "Case Selection in A Disconfirmatory Case Study" in *The Information Systems Research Challenge -- Volume 1: Qualitative Research Methods* (J. Cash and P. Lawrence, eds), Harvard Business School Publishing Division, Boston, MA, 20-26.
- Markus, M.L. (1994) "Electronic Mail as the Medium of Managerial Choice," *Organization Science*, 5 (4), 502-527.
- McKeen, J. and Guimaraes, T. (1997) "Successful Strategies for User Participation in Systems Development." *Journal of Management Information Systems*, 14 (2), 133-150.

- Mukhopadhyay, T., Kekre, S. and Kalathur, S. (1995) "Business Value of Information Technology: A Study of Electronic Data Interchange." *MIS Quarterly*, 19 (2), 137-156.
- Nagel, E. (1979) *The Structure of Science*, Hackett Publishing, Indianapolis, IN.
- Nandhakumar, J., and Jones, M. (1997) "Designing in the Dark: the Changing User-Developer Relationship in Information Systems Development" in *Proceedings of the Eighteenth International Conference on Information Systems* (Kumar, K. and DeGross, J., eds.), 75-87.
- Neter, J., Wasserman, W., and Whitmore, G. (1993) *Applied Statistics*, Fourth Edition, Allyn and Bacon.
- Newman, M. and Sabherwal, R. (1996) "Determinants of Commitment to Information Systems Development: A Longitudinal Investigation." *MIS Quarterly*, 20 (1), 23-54.
- Ngwenyama, O. and Lee, A. (1997) "Communication Richness in Electronic Mail: Critical Social Theory and the Contextuality of Meaning." *MIS Quarterly*, 21 (2), 145-167.
- Popper, K. (1968a) *Conjectures and Refutations: the Growth of Scientific Knowledge*, Harper Torchbooks, NY.
- Popper, K.. (1968b) *The Logic of Scientific Discovery*, Harper Torchbooks, NY.
- Robey, D., and Sahay, S. (1996) "Transforming Work through Information Technology: A Comparative Case Study of Geographic Information Systems in County Government." *Information Systems Research*, 7 (1), 93-110.
- Rosenberg, A. (1993) "Hume and the Philosophy of Science" in *The Cambridge Companion to Hume* (Norton, D., ed.), Cambridge University Press, New York, 64-89.
- Schultze, U. (2000) "A Confessional Account of an Ethnography about Knowledge Work." *MIS Quarterly*, 24 (1), 3-41.
- Smith, H., Milberg, S., and Burke, S. (1996) "Information Privacy: Measuring Individuals' Concerns about Organizational Practices." *MIS Quarterly*, 20 (2), pp. 167-196.
- Stein, E., and Vandenbosch, B. (1996) "Organizational Learning During Advanced System Development: Opportunities and Obstacles." *Journal of Management Information Systems*, 13 (2), 115-136.
- Straub, D. (1989) "Validating Instruments in MIS Research." *MIS Quarterly*, 13 (2), 147-169.

- Strauss, A., and Corbin, J. (1998) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Second Edition, Sage Publications, Thousand Oaks, CA.
- Stuchfield, N, and Weber, B. (1992) "Modeling the Profitability of Customer Relationships: Development and Impact of Barclays de Zoete Wedd's BEATRICE." *Journal of Management Information Systems*, 9 (2), 53-76.
- Tyran, C., Dennis, A., and Vogel, D. (1992) "The Application of Electronic Meeting Technology to Support." *MIS Quarterly*, 16 (3), 313-334.
- Vandenbosch, B. and Ginzberg, M., (1996) "Lotus Notes and Collaboration: Plus Ça Change..." *Journal of Management Information Systems*, 13 (3), 65-81.
- Walsham, G. (1993) *Interpreting Information Systems in Organizations*, Wiley, New York.
- Walsham, G. (1995) "Interpretive Case Studies in IS Research: Nature and Method." *European Journal of Information Systems*, 4, 74-81.
- Walsham, G. and Sahay, S. (1999) "GIS for District-Level Administration in India: Problems and Opportunities," *MIS Quarterly*, 23 (1), 39-65.
- Wonnacott, T. and Wonnacott, R. (1990) *Introductory Statistics for Business and Economics*, Fourth Edition, Wiley, NY.
- Yin, R. (1984) *Case Study Research, Design and Methods*, Sage Publications, Beverly Hills, CA.
- Yin, R. (1994) *Case Study Research, Design and Methods*, Second Edition, Sage Publications, Thousand Oaks, CA.