9-2012

Generalization and induction: Misconceptions, clarifications and a classification of induction

Eric W. K. TSANG
University of Texas at Dallas

John N. WILLIAMS
Singapore Management University, johnwilliams@smu.edu.sg

DOI: https://doi.org/10.2307/41703478

MIS Quarterly

ISSUES AND OPINIONS

Generalization and Induction: Misconceptions, Clarifications, and a Classification of Induction[1]

Eric W. K. Tsang
School of Management, University of Texas at Dallas, 800 W. Campbell Road,
Richardson, TX 75083-0688 U.S.A. {ewktsang@utdallas.edu}

John N. Williams
School of Social Sciences, Singapore Management University, 90 Stamford Road,
Singapore 178903 SINGAPORE {johnwilliams@smu.edu.sg}

In "Generalizing Generalizability in Information Systems Research," Lee and Baskerville (2003) try to clarify generalization and classify it into four types. Unfortunately, their account is problematic. We propose repairs. Central among these is our balance-of-evidence argument that we should adopt the view that Hume's problem of induction has a solution, even if we do not know what it is. We build upon this by proposing an alternative classification of induction. There are five types of generalization: (1) theoretical, (2) within-population, (3) cross-population, (4) contextual, and (5) temporal, with theoretical generalization being across the empirical and theoretical levels and the rest within the empirical level. Our classification also includes two kinds of inductive reasoning that do not belong to the domain of generalization. We then discuss the implications of our classification for information systems research.

Keywords: Research methodology, generalization, generalizability, induction, deduction, statistical generalization, statistical syllogism, inductive analogy, Hume's problem of induction

Introduction

Induction is essential for any substantial scientific endeavor. Yet there are misconceptions about induction and, in particular, about generalization in information systems research. Lee and Baskerville (2003) is the major treatment of generalization in the IS literature. They warn us that uncritical application of statistical sampling-based generalizability as the norm for all generalizability may result in inappropriate evaluations of many studies. They attempt to "clarify the concept of generalizability by critically examining its nature, illustrating its use and misuse, and presenting a framework for classifying its different forms" (p. 221). They are centrally concerned with "statistical generalization." This occurs when a researcher observes a characteristic of a sample of a population then infers that the population itself has that characteristic, within a margin of error. Lee and Baskerville claim that "its uncritical application as the norm for all generalizability can lead to an improper assessment of the generalizability of many research studies" (p. 221).

They also invoke Hume's problem of induction (originally published in 1748) to support their commitment to the claim that statistical generalization is unjustifiable.[2] Nonetheless

[1]Detmar Straub was the accepting senior editor for this paper.

[2]Lee and Baskerville say that "we affirm the legitimacy of statistical generalizability, where we emphasize that it refers to the generalizability of one random sample to other random samples that would result from applying the same sampling procedure to the same population or the generalizability of sample points to a sample estimate" (p. 228) as opposed to "the *generalizability of the sample estimate to the corresponding population characteristic*" (p. 226). They endorse "Hume's truism" that "*induction or generalization is never fully justified logically*" (p. 225), and add that "the latter generalizability, according to Hume's truism, cannot be improved (*or, for that matter, even established*) by increasing the sample size" (p. 226, emphasis added).

Table 1. Examples of Confusion Caused by Lee and Baskerville

Confusion	Example
Conflating induction with deduction	When discussing the advantages and possible difficulties of professionally qualified doctoral students, Klein and Rowe (2008) adopt Lee and Baskerville's classification. In their Table 2, which is adapted from Lee and Baskerville's Figure 5, they follow Lee and Baskerville in calling Type TE generalizability deduction. But generalization is a form of induction, not deduction.
Misunderstanding the nature of statistical generalization	Olsson et al. (2008, p. 265) state that Lee and Baskerville "present an overarching frame-work that proposes four distinct categories of generalizing, only one of which corresponds to statistical sampling-based generalization." Statistical generalization is included by Lee and Baskerville under Type EE generalizability which, as we will show, is problematic. In con-trast, statistical generalization is properly typified in our classification as within-population generalization.
Misunderstanding the relationship between theory and generalization	When discussing the implications of their action research results, Lindgren et al. (2004, p. 466) highlight Lee and Baskerville's point that "a theory may *never* be generalized to a setting where it has not yet been empirically tested and confirmed" (p. 241). As we will show, this claim is inconsistent with the meaning of theory.
Conflating empirical testing with generalization	Venkatesh and Ramesh (2006, p. 184) say that "Lee and Baskerville (2003) highlight the importance of establishing generalizability of a theoretical concept to different settings" (i.e., Type TE generalizability) and use Type TE generalizability to justify the contribution of their two empirical studies. But Type TE generalizability is empirical testing, not generalization.
Conflating empirical testing with generalization	When discussing the limitations of their study, Ragu-Nathan et al. (2008, p. 430) note that "replicating this study in other sectors, such as health care, for example, would lead to empirical generalization of these theoretical relationships (Lee and Baskerville 2003)." By "empirical generalization" they appear to mean Type TE generalizability. But replicating their studies in other sectors is testing, not generalizing, the theoretical relationships.

they claim that despite Hume's problem of induction, "a larger sample size does increase generalizability, but it is the generalizability of a sample to other samples, not to the population" (p. 227). Then they identify four types of generalization. These are empirical statement to empirical statement (EE), empirical statement to theory (ET), theory to empirical statement (TE), and theory to theory (TT). Lee and Baskerville's pioneering paper has proven to be influential. Alas it is not entirely unproblematic. In Table 1, we give some examples of the confusion caused by their treatment.

To grapple with Hume's problem of induction, we must first understand induction. An inference is *inductive* when and only when it is an inference from matters of fact that we have observed to those we have not (*Cambridge Dictionary of Philosophy* 1999, p. 745). Hume's problem of induction is his argument that the use of induction—including statistical generalization—is unjustifiable. Lee and Baskerville are committed to endorsing this and so must deny that an increase in sample size increases the probability that the population has the characteristic found in the sample. This is because any statistical generalization, however large the sample, is still as much an inductive inference as the case in which the sample consists of just one member of the population, and so is equally unjustified.

In the next section we explain Hume's problem of induction. Then we show that Lee and Baskerville's claim—that Hume's problem of induction imposes "no prohibition of the conclusion that an increase in sample size leads to an increase in the generalizability of one sample to other samples that the same sampling procedure would produce" (p. 228)—does not hold up to close scrutiny. In Appendix A, we give a balance-of-evidence argument that it is reasonable to assume that there is some solution to Hume's problem, even if we do not know what the solution is. In the following two sections we make objections to Lee and Baskerville's account of generalization as well as their classification of it into four types. We offer repairs of these problems. On this basis of discussion we propose a classification of induction, within which we distinguish five types of generalization. These are (1) theoretical, (2) within-population, (3) cross-population, (4) contextual, and (5) temporal, with theoretical generalization being across the empirical and theoretical levels and the rest within the empirical level. Our classification also includes two kinds of inductive reasoning that do not belong to the domain of generalization. Our classification of induction distinguishes it from generalization, and so rectifies conceptual confusion inherited from Lee and Baskerville's treatment of generalization. To further explain our objections to Lee and Baskerville, we discuss the implications of our classification of induction for IS research before ending with some brief concluding remarks. We also summarize the key points of Lee and Baskerville's paper in Appendix B.

Hume's Problem of Induction

To understand Hume's problem of induction we need to briefly review the terms of logic within which it is framed. An inference is *deductive* when and only when it is supposed to be *valid*; in other words, if its premises are true, then its conclusion must be true (Copi and Cohen 1990; Hurley 2003). Otherwise it is *invalid*. As we noted, it is *inductive* when and only when it is an inference from observed matters of fact to unobserved matters of fact. Deductive inferences may establish conclusions with certainty, whereas inductive inference may only establish them with high probability (Copi and Cohen 1990). An inductive inference is *strong* to the degree that its conclusion is probably true, given the truth of its premises (Hurley 2003).

Can we justify any inductive inference? If the answer is yes, then we may justify our belief that induction is reliable and so justify our use of induction. But Hume argues that no justification is possible. Take the following:

A P1 In all of countless previous cases without exception, objects that are terrestrial have been observed to obey the law of gravity.
  C All terrestrial objects will be observed to obey the law of gravity.

Plainly A is invalid because there is no contradiction in supposing that the law of gravity will suddenly cease operating. We could make it valid by adding an extra premise P2:

B P1 In all of countless previous cases without exception, objects that are terrestrial have been observed to obey the law of gravity.
  P2 If, in all of countless previous cases without exception, objects that are terrestrial have been observed to obey the law of gravity, then all terrestrial objects will be observed to obey the law of gravity.
  C All terrestrial objects will be observed to obey the law of gravity.

Now the problem is in justifying P2. P2 assumes that nature is uniform; in other words, that laws of nature will always continue to operate. But what reason is there to accept this? Once again, there is no contradiction in supposing that laws of nature, such as the law of gravity, will cease operating.

Why can't we just say an inference such as A is strong? Then we are making the following inference:

C P1 In past experience, all terrestrial objects have been observed to obey the law of gravity.
  C All terrestrial objects to be observed will *probably* obey the law of gravity.

This does not help because C is invalid. To make it valid, we must add a second premise P2:

D P1 In past experience, all terrestrial objects have been observed to obey the law of gravity.
  P2 If, in past experience, all terrestrial objects have been observed to obey the law of gravity, then all terrestrial objects to be observed will probably obey the law of gravity.
  C All terrestrial objects to be observed will probably obey the law of gravity.

But there is no reason to accept P2. Why should we think that scientific laws have a propensity to resemble the past in significant ways? One's answer cannot be that they have always had this propensity, for that would be to use induction, which is what needs justification.

We could point out that although inductive inferences with true premises are not guaranteed to have true conclusions, they have mostly had true conclusions in the past. But since we are interested in justifying our *continuing* use of induction, we would have to argue:

E P1 Past uses of inference A have given us conclusions that were mostly true.
  C Future uses of inference A will give us conclusions that are mostly true.

But E is itself an *inductive* inference. If the law of gravity stops operating forever, E will have a true premise and a false conclusion. We are back where we started.

In sum, we have failed to provide a deductive justification of induction. We have failed to provide an inductive justification of induction, since we cannot assume what we are trying to justify. Therefore, assuming that justification is either deductive or inductive, there is no justification of induction.

Given this dismaying result, it is a waste of time to use any inductive inference—including statistical generalization—in scientific research. It is equally a waste of time to use induction to formulate a general theory on the basis of a set of particular observations. It is even futile to report the single observation of a particular phenomenon. This is because once we have finished observing the phenomenon and started to report it, we need a reason to think that what we have started to report remains anything like what we have just observed. This reason must be inductive. We do not even have any reason for the commonsensical belief that the ground won't swallow us up at the next step. The upshot is that science—and commonsense as well—is a waste of time.[3]

However, as we indicate in Appendix A, the history of the philosophy of science is replete with attempts to solve the problem. Strawson (1952) argues—defensibly, so we try to show—that the very question of justification is misconceived and that at best the question amounts to an inappropriate imposition of the standards of deduction upon induction. The logical empiricist Reichenbach (1938) argues for the pragmatic conclusion that from a decision-theoretic perspective of maximizing utility, using induction is our best possible bet. It might be argued that Hume himself has a nonlogical justification in mind.[4]

Have Lee and Baskerville Escaped Hume's Problem of Induction?

Lee and Baskerville claim that although Hume's problem of induction cannot be solved, and that therefore no inductive inference can be logically justified, nonetheless "a larger sample size does increase generalizability, but it is the generalizability of a sample to other samples, not to the population" (p. 227). They also make their claim in a different way: "there is no prohibition of the conclusion that an increase in sample size leads to an increase in the generalizability of one sample to other samples that the same sampling procedure would produce" (p. 228). Here *prohibition* is elliptical for "prohibition by the insolubility of Hume's problem of induction."

By now it should be obvious why this claim does not hold up, given that Hume's problem of induction cannot be solved: to generalize from observation of a characteristic of a sample to a conclusion about that characteristic in other samples that we have not observed involves an inference from what we *have observed* to what we have *not observed*. If Hume is correct, this is not justified. In fact, if he is correct, we are not even logically justified in reporting the result of *any* observation, once we have made it. This is because we need a reason to think that once we have finished observing a phenomenon and started reporting, the phenomenon still resembles what we have just observed. For example, once we have finished observing that litmus paper has turned red, we need to assume that it remains red once we have shifted our attention away from it. Hume would argue that we have no logical justification for this assumption. This last point is related to what we call "temporal generalization" in our classification of induction. In our opinion, Lee and Baskerville have not considered the serious implications of Hume's problem of induction. Its insolubility means the end of science.

Moreover, they are committed in the following two ways to claiming that Hume's problem of induction is insoluble. First, they say that they will "focus on Hume's truism, which calls attention to an irremediable problem in induction" (p. 224). They endorse Rosenberg's "succinct description of Hume's truism" (p. 225), which is that induction may be justified neither deductively nor inductively, and that "accordingly, claims that transcend available data, in particular predictions and general laws, remain unwarranted" (Rosenberg 1993, p. 75). They call this "the problem of induction" (p. 224). It follows that they must say that the problem of induction, namely that induction is not justified or warranted in *any sense*, is insoluble, or as they say "irremediable."

Second, what makes them assert that "a theory generalized from the empirical descriptions in a particular case study has no generalizability beyond the given case" (p. 236) is the putative insolubility of Hume's problem of induction. Their view is also clearly reflected by their saying that "as a consequence of Hume's truism, a theory may *never* be generalized to a setting where it has not yet been empirically tested and confirmed" (p. 241).

In Appendix A, we argue for the verdict that there is some solution to Hume's problem, even if we do not know what the

---

[3]It is obviously unfair to object that this conclusion contradicts commonsense unless we have already faulted Hume's argument. It might be objected instead that Hume himself believed in the value of induction—both commonsensical and scientific. Fortunately we need not enter into textual exegesis to decide on the truth of this claim, because even if it is true, all that it shows is that Hume himself did not see the serious implications of his own problem. What Hume certainly thought is that our use of induction is a psychologically engrained habit—as he puts it, a "custom"—that we cannot shed. We cannot refrain from causal inference any more than we can refrain from breathing. This might be offered as a nonlogical justification of induction. But a skeptic could reply that this means that human nature dooms us to irrationality, because we cannot help that our thought processes run in a direction that is unjustified.

[4]In terms of custom as discussed in the preceding note.

solution is. This verdict has practical benefits: scientists may continue to trust the standard view that inductive inference is strong provided we have a big enough sample size relative to the population concerned when inferring from sample characteristics to population characteristics. It also removes what Lee and Baskerville think is an obstacle to the helpful and orthodox view held by researchers that an increase in the size of a sample strengthens the inference to the conclusion about the population from which it is drawn.

This view coheres with commonsense. Suppose that a random sample of 100 companies in Tokyo shows that 40 percent of them have the position of chief information officer. Another random sample of 1,000 companies in Tokyo also shows that 40 percent of them have the position of chief information officer. It hardly needs to be argued that the inference to the conclusion that about 40 percent of all companies in Tokyo have the position of chief information officer is stronger in the second case.

Our verdict also allows us to reach a theory by generalizing it from observations, although we may conjecture it as well. Another benefit of endorsing the verdict is that there is no longer an obstacle to producing a classification of generalization—and therefore no longer an obstacle to Lee and Baskerville's own classification.

We have shown that Lee and Baskerville do not appreciate the serious implications of Hume's problem of induction. We remedy this in Appendix A with a balance-of-evidence argument for the verdict that there is some solution to the problem. In order to motivate our classification of induction, we now turn to Lee and Baskerville's discussion of generalization and their classification of it into four types. We will discuss its limitations and propose repairs.

Lee and Baskerville on Generalization

There are three problems with Lee and Baskerville's discussion of generalization:

(1) Their definition of it is not the sense used by researchers in natural or social sciences.

(2) Their definition of induction is too narrow.

(3) Their characterization of generalization contradicts their own definition of it.

We discuss each of these problems in turn, proposing repairs.

Lee and Baskerville's definition of generalization is not the sense used by researchers in natural or social sciences. Lee and Baskerville cite the *Oxford English Dictionary* (1989) when defining *generalize* as "to form general notions by abstraction from particular instances" (p. 221). But lexicographers often give different definitions. They list primary senses first, followed by others that may be less common or more specialized. For generalize, the *Oxford English Dictionary* starts with "reduce to general laws," followed by "infer (a conclusion, law, etc.) inductively from particulars," "draw general inferences from," and then Lee and Baskerville's favored "form general notions by abstraction from particular instances" followed by "bring into general use" (as in "this style of writing e-mails has become more generalized"). The American dictionary, *Merriam-Webster*, adds "spread or extend throughout the body" (as in "the cancer has become more generalized from the pancreas," or analogically, "the distrust of IT has generalized within the company from the HR department"). Which sense of generalize are we supposed to think of? The question matters because although there may be multiple senses of a term, there are wrong usages of words.[5]

Moreover some usages of generalize might be irrelevant to IS researchers. Certainly the *Merriam-Webster* sense seems peripheral. What of the sense that Lee and Baskerville single out? One might naturally think of the movement from observation of particular phenomena to something more general as terminating not in a *conclusion,* but rather the *formation of a concept* (in other words, a *notion*) of X, the possession of which is just one's ability to reliably distinguish Xs from non-Xs. One difference between generalization as *inductive inference* and generalization as *concept formation* is that in concept formation there is no conclusion that needs to be ascertained to be true or false. Another difference is that coming to a conclusion is not coming to form a concept, although of course one needs to have already acquired concepts in order to arrive at a conclusion, because one must have had those concepts that are embodied in the thought of the conclusion. We admit that cognitive psychologists may well be interested in concept formation. But the IS researchers for whom Lee and Baskerville write are primarily interested in inference. Moreover, Lee and Baskerville focus on Hume's problem of induction, so generalization as a form of induction must be the type of generalization with which they are concerned.[6] So Lee and Baskerville's definition of generalization

[5]For example, one may not sensibly use "right" to mean "wrong" anymore than Humpty Dumpty sensibly used "glory" to mean "a nice knock-down argument" (Carroll 1871, p. 45).

[6]We owe this sharp point to a reviewer of our paper.

is not the central sense of the term as it is used in practice by such researchers.

Our repair is to propose a more precise, inference-based definition. For any two statements *P* and *Q*, *Q* is *more general* than *P* when and only when the class of entities that *P* is about is a subset of the class of entities that *Q* is about. Thus "all metals expand when heated" is more general than "all liquid metals expand when heated." Then we may say that for any two statements, *P* and *Q*, *Q* is a *generalization* of *P* when and only when (1) *Q* is inferred from *P*, and (2) *Q* is more general than *P*. Thus,

P1 All the metals we experimented with today in the lab expanded when heated.
C All metals expand when heated.

is a generalization because C is more general than P1, from which it is inferred. This characterization of generalization, in the context of logic, is consistent with that of Copi and Cohen (1990) and Hurley (2003).

(2) Lee and Baskerville's definition of induction is too narrow. Lee and Baskerville then define *induction* as "a reasoning process that begins with statements of particulars and ends in a general statement" (p. 224). This definition is too narrow, a point well made by Copi and Cohen and by Hurley, because there are at least two forms of inductive inference that do *not* proceed from particular premises to a general conclusion, both of which may be usefully employed in scientific contexts. One of these is what logicians call *statistical syllogism* (Gensler 2001), for example:

P1 Nearly all senior IT managers in the UK have college degrees.
P2 Tom is a senior IT manager in the UK.
C Tom has a college degree.[7]

This inference is inductive, because it goes from observed matters of fact to unobserved matters of fact and so cannot hope to establish its conclusion with certainty, but only with high probability. Yet the inference goes from a *general* premise and a particular premise to a *particular* conclusion. So it is not a generalization.

Another form of inductive inference is what logicians call *inductive analogy* (Copi and Cohen 1990; Hurley 2003), for example:

P1 Peter's computer has the Windows Vista operating system and McAfee anti-virus software, and Peter never visits dubious websites or downloads files from e-mails of unfamiliar senders to his computer.
P2 Mary's computer has the Windows Vista operating system and McAfee anti-virus software, and Mary never visits dubious websites or downloads files from e-mails of unfamiliar senders to her computer.
P3 Peter's computer is virus-free.
C Mary's computer is virus-free.

This inference is also inductive, because it also goes from observed matters of fact to unobserved matters of fact. But the inference goes from *particular* premises to a *particular* conclusion. So this is not generalization either.

We have demonstrated that Lee and Baskerville's definition of induction is too narrow, especially for practical use by researchers. The result is that scholars will probably ignore some types of induction that may be usefully considered. Our repair is to say that an inference is inductive when and only when it goes from observed matters of fact to unobserved matters of fact. Not only is this consistent with standard usage, it also has the advantage that IS researchers may consider how to use these two important forms of induction.

(3) Lee and Baskerville's characterization of generalization contradicts their own definition of generalization. Lee and Baskerville assert that "the generalizability of an IS theory to different settings is important" (p. 221). Their idea of generalizing from theory (i.e., general notions) to settings (i.e., particular instances) simply contradicts their own adopted definition of generalization, which refers to forming general notions by abstraction from particular instances.

Our repair is to replace the term *generalizability* in their statement with *applicability*, which refers to how far a theory covers certain empirical phenomena. A theory is more general when its domain of applicability is wider, and accordingly, its predictive and explanatory power greater. Later in their article, Lee and Baskerville ask

> what might scientific researchers do so that they may recommend their theories for application in new

[7]An inference of the form

P1 Nearly all (or most or 67%) Fs are Gs.
P2 X is an F.
C X is a G.

has two premises and so is *ipso facto* a syllogism, albeit not a categorical syllogism. The syllogism is "statistical" in the sense that the first premise reports a precise statistic (67%) or a vague range of statistics. Thus "most Fs are Gs" means that more than 50% but less than 100% of Fs are Gs. "Nearly all Fs are Gs" means that less than 100% of Fs are Gs but significantly more than 50%—arguably more than 90%—of Fs are Gs. At the upper boundary of 100% the inference becomes deductively valid. The term *statistical syllogism* is standard in logic and philosophy of science.

> settings, where the theories have not yet been empirically tested and confirmed in those settings? (p. 240)

This time, in line with our critique, they use the proper term *application*. Our alternative characterization avoids all these problems.

Lee and Baskerville's Four Types of Generalization

Lee and Baskerville classify generalization into four types. Their classification is based on a distinction between empirical and theoretical statements:

> Empirical statements can refer to data, measurements, observations, or descriptions about empirical or real-world phenomena, while theoretical statements posit the existence of entities and relationships that cannot be directly observed, and hence can only be theorized (p. 232).

Based on this distinction, they classify generalization as empirical statement to empirical statement (EE), empirical statement to theory (ET), theory to empirical statement (TE), and theory to theory (TT).

This distinction faces two objections. First, their characterization of an empirical statement cannot serve as a definition because it is circular; to understand what makes a statement empirical, we must first understand what makes real-world phenomena empirical. Since *empirical* means "based on what is experienced" (*Cambridge International Dictionary of English*), our repair is to say that a statement is empirical when and only when it cannot be ascertained to be true or false without experience or observation. This definition coheres with *empirical statements* defined by *Merriam-Webster Dictionary* as those "capable of being verified or disproved by observation or experiment," since both observation and experiment involve experience. Even an anthropological interpretivist who produces a rich description of a particular Maasai in the Serengeti without generalizing this to conclusions about other Maasai would accept that a statement such as "Idi wears cowhide sandals and carries a wooden club" is empirical in the philosophically uncontentious way we have defined it.

Second, theoretical statements need not posit the existence of entities and relationships that cannot be directly observed. They may or may not posit them. Newton's laws of motion certainly constitute a theory, but they are about observable relations between perfectly ordinary, observable objects. Our repair is to say that a statement is theoretical when and only when it is a generalization that purports to predict and explain the phenomena to which it refers. This accommodates the fact that some, but not all, theoretical statements posit relationships between what is directly observable.

As we will now show, Lee and Baskerville go on to mischaracterize each of the so-called four types of generalization, even within their own terms. In particular, only what they call "Type ET generalizability" is really generalization.

Type EE Generalizability: Generalizing from Data to Description

There are three problems with Lee and Baskerville's discussion of this type. First, in order to illustrate the generalization of data to a measurement, Lee and Baskerville use the example of calculating the sample mean from sample points (pp. 233-234). But a sample estimate is *deduced* and not induced from sample points. To fix this problem, they should either drop the example from their position or use it to illustrate deduction instead.

Second, in stressing the importance of validating a measurement instrument, Lee and Baskerville state:

> In the situation where the measurement instrument has not been validated, the data collected from a research subject would lack generalizability to any valid measurement for that individual (p. 234).

This does not involve any sense of *generalize* that characterizes inference. If an instrument of measurement, such as a thermometer, has not been validated in the sense of having been checked for accuracy, it is uncertain whether the data generated by the instrument accurately indicate what is intended to be measured. This straightforward methodological point has nothing to do with generalizability. Of course Hume might add that even if it has been checked for accuracy (past readings of the thermometer having been observed to be accurate), that is no reason to think that it will continue to be so. Thus, since Lee and Baskerville think that Hume's problem of induction is insoluble, there is no point in classifying this type of generalization in the first place.

Third, Lee and Baskerville claim,

> From this particular set of field data, an ethnographer could generalize the description that, in the world of these officers, autonomy is indeed highly valued—so much so, in fact, that the officers will

> conjure up busywork to satisfy their sergeant, distract his attention, and thereby otherwise preserve their autonomy (p. 234).

The description in question is supposed to be an empirical statement because Type EE generalizability is about generalizing from empirical statements to other empirical statements. However, Lee and Baskerville state that "theoretical statements posit the existence of entities and relationships that cannot be directly observed, and hence can only be theorized" (p. 232). But surely autonomy and attention cannot be directly observed—at the very least not in the way in which we directly observe litmus paper turning red.[8] So the description is a theoretical statement, and should not be classified as Type EE generalizability. Another supposedly empirical statement, "autonomy is highly valued in the eyes of the officers" (p. 235), faces the same problem. One remedy is to adopt the definitions of empirical and theoretical statements we recommend above. According to these definitions, Lee and Baskerville's discussion is about generalizing from empirical data to theoretical descriptions. Thus, it belongs to Type ET generalizability.

Type ET Generalizability: Generalizing from Description to Theory

Based on Hume's argument, Lee and Baskerville state that "a theory generalized from the empirical descriptions in a particular case study has no generalizability beyond the given case" (p. 236). They reiterate this claim in a slightly different way: "a theory may *never* be generalized to a setting where it has not yet been empirically tested and confirmed" (p. 241). But a theory consists of general statements that extend beyond the empirical data from which the theory is developed. More formally, a theory is "a hypothetico-deductive system—that is, a system of hypotheses within which valid arguments (i.e., deductive chains) can be constructed" (Bunge 1996, p. 114). If a theory cannot be applied to phenomena beyond such data, it should not be called a theory in the first place (see Nagel 1979).

Type TE Generalizability: Generalizing from Theory to Description

There are two problems with Lee and Baskerville's discussion of this type, both of which are remediable. First, they stipulate that

> Type TE generalizability…involves generalizing from theoretical statements…to empirical statements (here, descriptions of what the practitioner can expect to observe in his specific organization if he were to apply the theory) (p. 237).

Although empirical statements may be general, such as "all big companies currently in Tokyo have management information systems departments," theoretical statements are, by their nature, necessarily general. Therefore, Lee and Baskerville's own definition of generalization, which has it that generalization consists in "generalizing *from* particular instances *to* general notions" (p. 232) rules out TE inferences as cases of generalization. So Lee and Baskerville contradict themselves within their own terms.

Second, Lee and Baskerville claim that "Type TE generalizability happens to be closely related to empirical testing" (p. 237). But Type TE generalizability just *is* empirical testing. It is related to generalization in the sense that if a theory's prediction is confirmed in settings that are different from the one where the theory has been tested and confirmed, this indicates that the results of the original setting might be generalized to other settings. Nevertheless, this concerns generalizing research results from one setting to others, not generalizing from theoretical statements to empirical statements. Our repair is to relabel what Lee and Baskerville call Type TE generalizability as *empirical testing* or *deductive prediction*. In fact, they approach this repair themselves by saying that Type TE generalizability "is actually deduction, not induction" (p. 241).

Type TT Generalizability: Generalizing from Concepts to Theory

There are again three problems with Lee and Baskerville's discussion of this type: the discussion is incoherent, internally inconsistent, and irrelevant to generalization in the true sense of the term.

First, Lee and Baskerville stipulate that "in Type TT generalizability, a researcher generalizes from theoretical propositions in the form of concepts…to the theoretical propositions that make up a theory" (p. 238). But this stipulation is incoherent, because concepts are not propositions. Even a theoretical concept such as the concept of an electron is not itself a proposition, theoretical or not. Understanding the theoretical proposition that all metals expand when heated requires possessing the concepts of metals and heat, but it does not follow that this proposition is a concept. If one acquires the concept of a metal then one acquires the ability to reliably distinguish metals from nonmetals, but one does not thereby acquire a theory.

[8] Nor can what the redness of litmus paper measures, namely acidity, be directly observed.

Second, Lee and Baskerville's own definition of generalization rules out generalizing from concepts to theories, because concepts are not particular instances. An electron is an instance of the concept of an electron, but it is not a concept, only an electron. Concepts are used to formulate a theory (Kaplan 1964), and cannot be generalized to a theory. Thus Lee and Baskerville's account is internally inconsistent.

Third, their discussion of the first form of Type TT generalization is in fact about inconsistent operationalization of constructs. The discussion of the second form is about constructing a theory from concepts extracted from the literature. But none of these discussions are related in the remotest sense to inference, let alone generalization. So their discussion is irrelevant to generalization in the true sense of the term. Given this, it is no wonder that Lee and Baskerville's discussion of Gefen and Straub's (1997) study does not clarify how we are supposed to generalize from concepts to theory. In fact, it may lead us into misunderstanding. One repair is to delete the reference to generalization and focus on discussing construct operationalization and theory construction.

This is not to say, however, that there is no connection between less general theories and more general ones. A fact may be *explained* by a theory and that theory also may be explained by a more general theory. For example, the fact that all liquid metals expand when heated may be explained by the theory that all metals expand when heated. The fact that all metals expand when heated may be explained by the more general theory that heating something raises its molecular kinetic energy. Thus less general theories may be deduced from more general ones. The *explanandum* may be seen as a theory or a fact. We may choose to call it a fact that all liquid metals expand when heated, but it is not an empirical fact, because it is impossible to observe every liquid metal—past, present, and future in any spatial location of the universe—expanding when heated. Hume, of course, might say that this means that we have no business calling it a fact.

The same points hold for social sciences. Consider, for instance, Van Maanen's (1983) point that policemen stake out bars frequented by drunk patrons, as cited in Lee and Baskerville's discussion of Type EE generalizability. The policemen's behavior—which is not itself a theory—can be explained by the theory that policemen stake out locations where armed robberies are likely and that bars where drunk patrons are found are such locations. This theory is in turn explained by a more general theory that officials in authority stake out locations where violent crime is likely and that places where drunks are found are such locations—because the less general theory may be deduced from the more general one. By contrast, no theory may be induced from another because induction is an inference from observed matters of fact to unobserved matters of fact, but while theories explain empirical observation, they are not observed matters of fact.

A Summary of Problems and Solutions

We have gone a long way in discussing the problems of Lee and Baskerville's paper and proposing repairs. Before presenting our classification of induction in the next section, we summarize these problems and repairs in Table 2.

A Classification of Induction

To classify induction, we first distinguish two levels of research activities, namely empirical and theoretical, which correspond broadly to Lee and Baskerville's distinction between empirical and theoretical statements. On the empirical level, researchers collect data through observation; on the theoretical level, researchers develop theories based on data collected on the empirical level, pure conjectures, or a combination of both. The classification is pictorially presented in Figure 1, where the arrows represent inferences. There are five basic types of generalization: (1) theoretical, (2) within-population, (3) cross-population, (4) contextual, and (5) temporal, with theoretical generalization being across the empirical and theoretical levels and the rest within the empirical level. Our classification also includes the two above-mentioned kinds of inductive reasoning that do not belong to the domain of generalization: statistical syllogism and inductive analogy. Both are within the empirical level.

The five types of generalization have their roots in the literature. To begin with, Gomm et al. (2000) and Sharp (1998) classify generalization into two main types: theoretical and empirical. Our classification further classifies empirical generalization into four types. Theoretical generalization corresponds to Yin's (2003) analytic generalization, in which "the investigator is striving to generalize a particular set of results to some broader theory" (p. 37), and embodies one of the five fundamental principles that scientists use in making generalizations (Shadish et al. 2002). Within-population generalization and cross-population generalization are concerned with where a sample is located—inside or outside—relative to the population at which generalization is targeted. A similar distinction between these two kinds of generalization can be found in the literature. For example, Maxwell (1992) distinguishes between internal and external

Table 2. A Summary of Solutions to the Major Problems in Lee and Baskerville		
Problem	**Evidence**	**Solution**
Hume's problem of induction		
Non-recognition of the serious implications of accepting that Hume's problem of induction is irremediable	" there is no prohibition of the conclusion that an increase in sample size leads to an increase in the generalizability of one sample to other samples that the same sampling procedure would produce" (p. 228)	Adopt a two-stage balance-of-evidence argument that it is reasonable to assume that there is some plausible solution to Hume's problem, even if we do not know which solution it is
Definitions of generalization and induction		
Definition of generalization is not consistent with the sense used by researchers in the natural and social sciences	" to form general notions by abstraction from particular instances" (p. 221)	Adopt an alternative definition: for any two statements *P* and *Q*, *Q* is a generalization of *P* when and only when 1) *Q* is inferred from *P*, and 2) *Q* is more general than *P*
Definition of induction is too narrow	" a reasoning process that begins with statements of particulars and ends in a general statement" (p. 224)	Adopt an alternative definition: an inference is inductive when and only when it goes from observed matters of fact to unobserved matters of fact
Statement about generalizability contradicts Lee and Baskerville's own definition of generalization	" the generalizability of an IS theory to different settings is important" (p. 221)	Replace "generalizability" in the statement by "applicability"
Definitions of empirical and theoretical statements		
Definition of empirical statements is circular	"Empirical statements can refer to data, measurements, observations, or descriptions about empirical or real-world phenomena" (p. 232)	Adopt an alternative definition: a statement is empirical when and only when it cannot be ascertained to be true or false without observation or experience
Definition of theoretical statements as necessarily positing the existence of entities and relationships that cannot be directly observed	" theoretical statements posit the existence of entities and relationships that cannot be directly observed, and hence can only be theorized" (p. 232)	Adopt an alternative definition: a statement is theoretical when and only when it is a generalization that purports to predict and explain the phenomena to which it refers
Type EE generalizability		
Example of deduction illustrates the generalization of data to a measurement	The example of calculating the sample mean from sample points on pp. 233-234	Drop the example or use it to illustrate deduction
Statement about validating measurement instrument does not involve any sense of generalization that characterizes inference	"In the situation where the measurement instrument has not been validated, the data collected from a research subject would lack generalizability to any valid measurement for that individual" (p. 234)	Delete the statement and its related discussion
Statement about field data contradicts Lee and Baskerville's own definitions of empirical and theoretical statements	"From this particular set of field data, an ethnographer could generalize the description that, in the world of these officers, autonomy is indeed highly valued — so much so, in fact, that the officers will conjure up busywork to satisfy their sergeant, distract his attention, and thereby otherwise preserve their autonomy" (p. 234)	Adopt the above-stated definitions of empirical and theoretical statements and place the statement and its related discussion under Type ET generalizability

Table 2. A Summary of Solutions to the Major Problems in Lee and Baskerville (Continued)

Problem	Evidence	Solution
Type ET generalizability		
Statement about "theories" generalized from case study descriptions entails that these are not theories	" a theory generalized from the empirical descriptions in a particular case study has no generalizability beyond the given case" (p. 236)	Delete the statement and its related discussion
Type TE generalizability		
Definition of Type TE generalizability contradicts Lee and Baskerville's own definition of generalization	"Type TE generalizability…involves generalizing from theoretical statements…to empirical statements (here, descriptions of what the practitioner can expect to observe in his specific organization if he were to apply the theory)" (p. 237)	Relabel Type TE generalizability as "empirical testing" or "deductive prediction" and revise the discussion accordingly
Nonrecognition that Type TE generalizability *is* empirical testing and not generalization	"Type TE generalizability happens to be closely related to empirical testing" (p. 237)	Relabel Type TE generalizability as "empirical testing" or "deductive prediction" and revise the discussion accordingly
Type TT generalizability		
Unintelligible and internally inconsistent stipulation	"In Type TT generalizability, a researcher generalizes from theoretical propositions in the form of concepts … to the theoretical propositions that make up a theory" (p. 238)	Delete the stipulation and its related discussion
Discussion of inconsistent operationalization of constructs is unrelated to generalization	Discussion of the first form of Type TT generalizability on p. 238	Delete the reference to generalization and focus on discussing the operationalization of constructs
Discussion of constructing a theory from concepts extracted from the literature is unrelated to generalization	Discussion of the other form of Type TT generalizability on p. 238	Delete the reference to generalization and focus on discussing the construction of theories

generalizability in qualitative research, corresponding to our within- and cross-population generalization, respectively. Similarly, Schofield (1990, p. 201) states "the aspect of external validity that has typically received the lion's share of attention in textbook and other treatments of the concept is generalizing to and across populations."[9] Contextual generalization and temporal generalization are derived from the insightful comment of Nagel (1979), discussed below, concerning the space–time constraint of social phenomena.

There are scattered discussions of the two types of generalization in the literature. Lucas (2003), for example, discusses generalizing across settings, corresponding to contextual generalization. Cronbach's (1975) argument that generalizations in the social sciences decay rather fast over time is related to temporal generalization.

For the two kinds of inductive reasoning, inductive analogy bears some resemblance to Stake and Trumbull's (1982) concept of naturalistic generalization. They argue that compared to quantitative studies, case studies are more epistemologically in harmony with the reader's experience and, thus, to that person, they form a natural base for generalization. A case often provides a thick description and a vicarious experi-

[9]The term *external validity* has been defined in various ways during the last three decades. For example, Cook and Campbell (1979, p. 37) defined it rather imprecisely as "the approximate validity with which we can infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings, and times." Later Cronbach (1982) developed an elaborate notional system that covers four dimensions: unit of analysis, treatment, outcome, and setting. Using the system, he distinguished between internal and external validity.

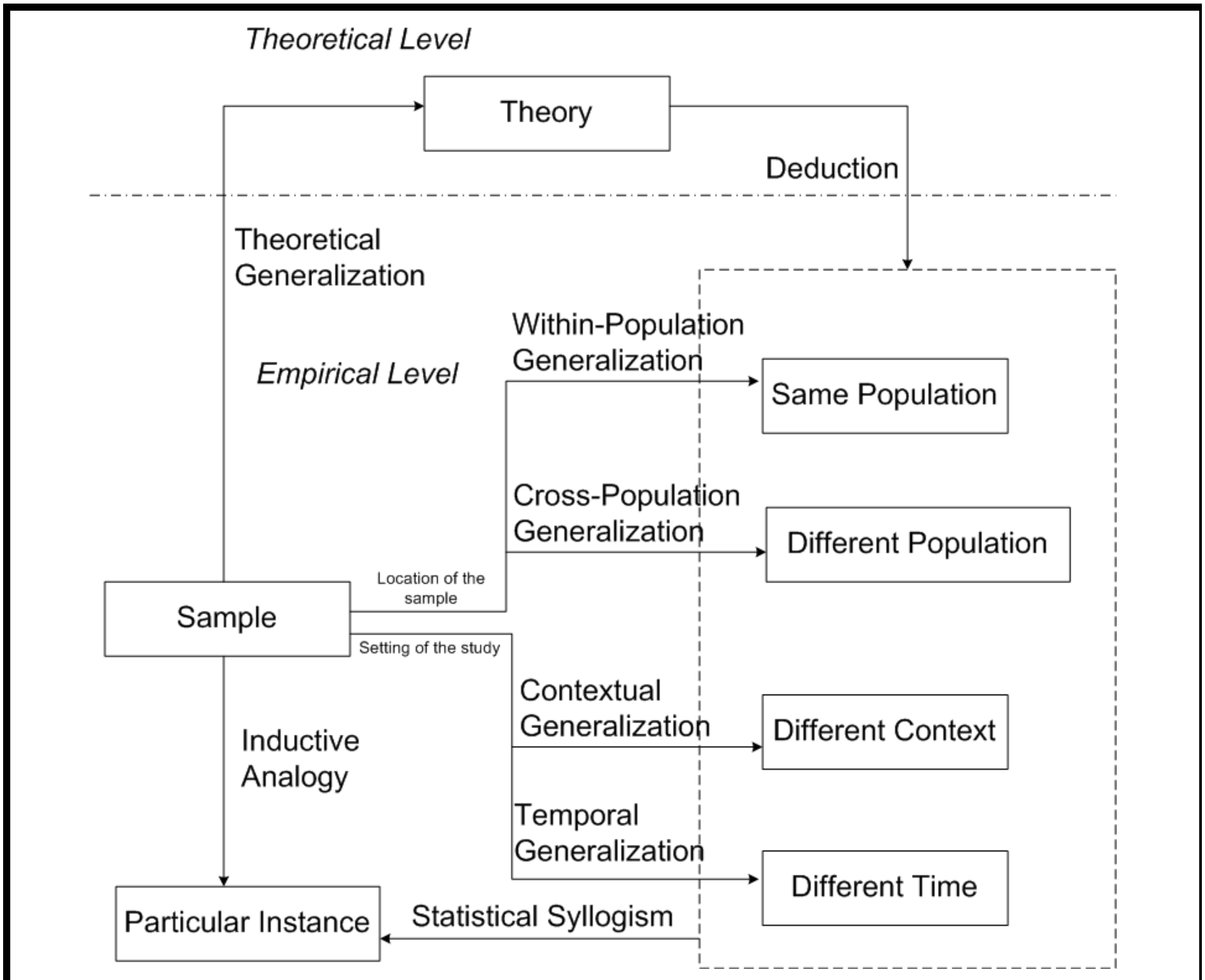

Figure 1. Types of Induction

ential account. The reader should be able to determine if and how far the case can be used to understand a new setting. Naturalistic generalization is a kind of tacit recognition of the similarities and differences between a case and a new setting, drawing on the personal experience and knowledge of the reader. In this sense, naturalistic generalization is somewhat similar to inductive analogy. Yet, as we argue above, this kind of reasoning is not generalization and the term *generalization* is thus misused by Stake and Trumbull. In the following discussion, we follow Lee and Baskerville and use Gefen and Straub's (1997) study to illustrate our classification. For easy reference, we list the definitions of each type of induction in Table 3.

Five Types of Generalization

Theoretical generalization, which is similar to Lee and Baskerville's Type ET generalizability, consists of generalizing from research findings to theories. Deriving theories from data collected in the empirical world is a traditional view of scientific research (Chalmers 1999). Gefen and Straub's study is a good example of this. Their findings suggest that the same mode of communication may be perceived differently by the sexes, and so they propose to extend the technology acceptance model by including the effects of gender. In other words, they generalize their results to a revised theoretical model.

Table 3. Definition of Types of Induction

Type of Induction	Definition
Theoretical generalization	Generalizing from research findings to theories
Within-population generalization	Generalizing from the characteristics of a sample to those of the corresponding population
Cross-population generalization	Generalizing from a sample in one population to members of another population, with both populations existing in a similar context and a similar period of time
Contextual generalization	Generalizing from a sample in one population to members of another population, with both populations existing in significantly different contexts but within a similar period of time
Temporal generalization	Generalizing from a sample in one population at one point in time to members of the same or a different population at another point in time, assuming that the context remains more or less the same.
Statistical syllogism	An inference of the form P1 N % of Fs are Gs. P2 X is an F. C X is a G. where "N" denotes a precise statistic or a vague range of statistics as in "Most" or "Nearly all."
Inductive analogy	An inference of the form P1 X has properties a, b, c … and z P2 Y has properties a, b, c … C Y has property z

Within-population generalization, which is equivalent to statistical generalization, consists of generalizing from the characteristics of a sample to those of the corresponding population. We prefer to use the term *within-population generalization* instead of *statistical generalization* in order to contrast this type of generalization with cross-population generalization. To perform within-population generalization, researchers need to first clearly specify the intended population in which their study is conducted, and then follow a sampling procedure that allows statistical inference from sample characteristics to population characteristics. Unfortunately, difficulties of research access often prevent researchers from discussing how far their findings are statistically generalizable. That said, researchers should at least clearly describe their sampling procedure. Gefen and Straub's study has some room for improvement in this respect. They distributed questionnaires to users of the e-mail system in one U.S., one Swiss, and one Japanese airline, without mentioning the number of employees in each airline. The only piece of information about the sampling frame is this sentence: "The sample included workers across managerial, professional, and technical ranks" (p. 395). That is, employees across these three ranks in each airline constituted a population. Although they report the response rate in each airline, they do not report the sample size drawn from each airline. As mentioned, a larger sample size will usually increase the within-population generalizability of the results of a study.

Cross-population generalization consists in generalizing from a sample in one population to members of another population, with both populations existing in a similar context and a similar period of time. Assume that Gefen and Straub had broken up their data into three parts, each of which contained data collected from one airline. In this case, cross-population generalization might consist in generalizing the results of the American airline (in which Gefen and Straub collected data) to another American airline that existed at the time of data collection. The contexts in which the two firms operate are similar in the sense that they are in the same industry and the same country. Note that a different way of defining a population will give rise to a different kind of cross-population generalization. Suppose that Gefen and Straub had conducted their study in only one of the three airlines, and had drawn a sample from the male employees of the airline only. The population would be the collection of male employees and one kind of cross-population generalization would consist of generalizing the findings to the population of female employees within that airline. As the actual findings of Gefen and Straub's study indicate that women and men differ in their perceived social presence and perceived usefulness of e-mail, the cross-population generalizability from male to female employees would be weak.

In comparing the social sciences with the natural sciences, Nagel (1979, p 459) observes that, "unlike the laws of physics and chemistry, generalizations in the social sciences therefore have at best only a severely restricted scope, limited to social phenomena occurring during a relatively brief historical epoch within special institutional settings." This observation implies that social scientists have to investigate whether their research findings collected in one space–time setting are generalizable to other significantly different space–time settings; in other words, whether these findings are contextually and temporally generalizable.

In *contextual generalization*, researchers generalize from a sample in one population to members of another population, with both populations existing in significantly different contexts but within a similar period of time. As the domain of management research in general, and IS research in particular, is becoming more international, the applicability of theories developed in one national context to other national contexts has been seriously questioned (Hofstede 1993; Rosenzweig 1994). Some scholars call for contextualization as "a way of approaching research where knowledge of the settings to be studied is brought to bear in design and implementation decisions" (Rousseau and Fried 2001, p. 6). By paying greater attention to the context in which a study is conducted, researchers are in a better position to assess the contextual generalizability of their findings. For example, Gefen and Straub specifically included the effect of cultural differences in their structural equations, and found that the effect was significant. This suggests that the generalizability of their findings to other cultural contexts would be limited. (We will discuss this issue in more detail below.) There are other types of context in addition to cultural or national context. Gefen and Straub conducted their study in the airline industry. The generalizability of their findings to another industry depends on whether the new industrial context significantly affects people's perception and usage of e-mail.

Temporal generalization consists in generalizing from a sample in one population at one point in time to members of the same or a different population at another point in time, assuming that the context remains more or less the same. Temporal generalization is the least discussed in the literature. Yet, it is implicitly present in any form of generalization in the sense that the act of generalizing and the findings to be generalized necessarily occur at different times. Temporal generalization is a more important issue in the social than the natural sciences. Nature is unaffected by scientific research. The fact that the earth revolves around the sun will not be changed by a theory proposing otherwise. However, activities of management researchers may change the beliefs and practices of managers and thus undermine the stability of the phenomena investigated (Numagami 1998). Self-fulfilling prophecy is a significant issue. For example, managers who are educated about the logic of transaction cost economics tend to impose controls for curbing opportunism. However, such controls may have the consequence of encouraging more difficult-to-detect opportunistic behavior (Ghoshal and Moran 1996). In short, managers may alter their behavior on the basis of the knowledge created by researchers (Knights 1992), and temporal generalization is a very important concern when one tries to generalize the findings of a study conducted in a distant past to a current situation. Gefen and Straub's study was done more than 10 years ago. Can their findings be generalized to today's airlines in the United States? Note that the airline industry in the United States has undergone significant changes since the September 11 terrorist attack. The much tighter security control in the industry may affect how e-mail systems are structured and how e-mails are used and monitored.

Generalization within the empirical level often involves a combination of some of the four basic types discussed above. For example, if researchers try to generalize Gefen and Straub's findings to today's airlines in Germany, this involves across-population, contextual, and temporal generalization. A major problem of this kind of generalization is that if the generalization is subsequently found to be wrong, it may be difficult to identify whether the failure is due to differences in population or contextual or temporal factors.

As indicated by Figure 1, theory, or the collection of theoretical knowledge, may be able to provide some useful information about the four types of generalization within the empirical level through deductive inference. Gefen and Straub, for instance, draw heavily on Hofstede's (1980) cultural dimension of masculinity–femininity in their study. The three countries in which they conducted their fieldwork, namely Japan, Switzerland, and the United States are high in masculinity. In these countries, gender roles are clearly defined and nurtured. So the gender effects that Gefen and Straub found in their study will probably be less generalizable to countries that are low in masculinity, such as Norway and Sweden. In short, knowledge of the cultural dimension of masculinity–femininity throws light on the contextual generalizability of their findings.

Statistical Syllogism and Inductive Analogy

After covering the five types of generalization, we proceed to discuss statistical syllogism and inductive analogy. Figure 1 shows that in *statistical syllogism*, researchers draw a conclusion about a particular instance from a generalization. Gefen and Straub's findings support the proposition that women perceive a higher social presence and usefulness of e-mail than men. Suppose we are interested in the differences in the perception of e-mail between a particular pair of male and female employees in the American airline in which Gefen and Straub collected data. We may use statistical syllogism and conclude that the female employee is likely to perceive a higher social presence and usefulness of e-mail than her male counterpart. Lee and Baskerville's Type TE generalizability is related to statistical syllogism in the sense that a generalization (i.e., theoretical statement) may be used to draw a conclusion about a particular instance (i.e., empirical statement).

In *inductive analogy*, researchers proceed from particular premises to a conclusion about a particular instance. Assume that Gefen and Straub had run their analysis separately on the data collected from the American airline, and suppose we are interested in the gender differences in the perception of e-mail among a group of employees drawn from another American airline. If there are substantial relevant similarities between this airline and Gefen and Straub's as well as between this group of employees and Gefen and Straub's sample, we may conclude that their results probably also reflect the characteristics of the group. Should Gefen and Straub provide more information about the airline and its employees that they studied, this would facilitate inductive analogy. Lee and Baskerville's Type EE generalizability is related to inductive analogy in the sense that a finding obtained from one setting (i.e., empirical statement) may be used to draw a conclusion about a new but similar setting (i.e., empirical statement). Unfortunately, when discussing Type EE generalizability, Lee and Baskerville have ruled out inductive analogy by claiming that "no descriptive statement (whether quantitative or qualitative) is generalizable beyond the domain that the researcher has actually observed" (p. 235). Finally, their Type TT generalizability is not a form of induction at all.

Implications for IS Research

Our classification helps IS researchers articulate the limitations of their studies. Since there are distinct types of generalization, when IS researchers discuss the generalization of their findings, it is important that they specify the kind of generalization they have in mind. However, this is often not the case. A typical example is: "As a small sample-size, single case study, generalizability cannot be assessed" (Majchrzak et al. 2000, p. 594). A pertinent question is: Generalize to what? Our classification highlights the need for researchers to think about the kind of generalization they refer to when they discuss the generalizability of their findings. This would help researchers clearly specify the limitations of their studies and point to concrete future research directions that may address these limitations. Gefen and Straub's discussion is an excellent illustration:

> From the standpoint of external validity, the study gathered data from three firms in one industry across three countries, which, *per force*, limits the generality of the results. It may well be, for instance, that there is a systematic bias in the airline industry that restricts our ability to generalize to other industries. Moreover, knowledge workers in managerial, professional, or technical positions in Japan tend to be overwhelmingly male, which very likely explains a relatively smaller number of Japanese women in our sample (p. 397).

Their reference to the airline industry is about generalizing across industrial contexts. Their highlighting of the idiosyncrasy of Japan seems to indicate their concern about the contextual generalizability of their Japanese case to companies in other countries. Further studies may be designed to test the generalizability of their findings along these directions.

Among the five types of generalization in our classification, temporal generalization is the most neglected. Researchers are usually aware of it when there is some significant structural change in the context of the original study. For example, researchers would hesitate to generalize the results of a study that investigated the use of IT in factories during the Soviet Union era to present-day Russia. However, temporal generalization can also be an issue when progressive change is gradual. A study of people's attitudes toward online shopping conducted 10 years ago is probably outdated as online shopping has become more popular and secure during the last 10 years. In brief, IS researchers should pay more attention to temporal generalization when citing previous studies.

Our classification also helps to clarify the debate about whether qualitative studies are less generalizable than their quantitative counterparts. We agree with Lee and Baskerville's view that within-population generalization (what they call statistical generalizability) is just one form of generalization and that qualitative research should not be evaluated based on this form of generalization alone. The findings of a case study are surely less generalizable to the population on which it is based, than the findings of a large-scale random-sample survey. But even in this case, the generalizability of survey results depends a great deal on whether a probability sampling method is employed to construct a representative sample. Various practical constraints, such as the inability to include every member of a population in the sampling procedure, often renders probability sampling infeasible in surveys. Thus the within-population generalizability of many survey results is questionable, an important issue that has been neglected by those who challenge case studies on the grounds of generalizability.

For the other four types of generalization, there is simply no reason why survey results should be inherently more generalizable than case study results. For instance, there is no justification to believe that case study results are definitely less generalizable than survey results *across* populations. If the population to which the findings are generalized is in a very different context and the contextual factors would affect these findings, the findings of both the case study and the survey are equally ungeneralizable to that population.

On the other hand, case studies have an edge over quantitative studies in terms of theoretical generalization. Case studies seek to investigate phenomena in their contexts, not independently of it (Gibbert et al. 2008), and throw light on the specific contingent conditions under which the postulated mechanisms operate to generate the phenomena (Tsoukas 1989). Such a research process allows researchers to tease out ever-deepening layers of reality in the search for mechanisms and influential contingencies (Harrison and Easton 2004). The rich description of a case study may lead one to conjecture a theory that one may then try to falsify by deducing predictions from it. It is no wonder that Walton (1992, p. 129) makes this bold claim: "case studies are likely to produce the best theory."

Conclusion

We share Lee and Baskerville's view that generalization is an important methodological issue and that statistical, sampling-based generalization is just one type of generalization. We also appreciate their attempt to clarify the concept of generalization and to develop a taxonomy of generalizations. Yet, their attempt may be improved. They do not seem to appreciate the drastic fall-out from granting that Hume's problem of induction cannot be solved. We have given Lee and Baskerville a useful way out of this impasse, by arguing that we should accept, at least as a working hypothesis, that it has a solution.

We have clarified the concept of generalization by locating the major problems in Lee and Baskerville's article. A main reason for these deficiencies is that they have not properly understood Hume's problem of induction. We therefore provided a deeper discussion of Hume's problem and presented a balance-of-evidence argument for the verdict that there is some justification of induction. Finally, we proposed a classification of induction, which will help researchers systematically organize the discussion about the generalizability of their research results.

Generalization is not only an important but also a complicated methodological issue, deeply embedded in logic, science, and the essentially argumentative discipline of philosophy. It is no wonder that misconceptions abound in the literature. For this reason, pioneers like Lee and Baskerville should be respected for leading the topic into the IS community, despite their mistakes. Our essay is just a second step toward elucidating the difficult concept of generalization and classifying different types of generalization. It is a topic that deserves more attention from methodologists, whatever their discipline.

Acknowledgments

The authors thank Detmar Straub (editor-in-chief) and the reviewers for their valuable comments. The authors' names are in alphabetical order, and both authors contributed equally to the article.

References

Bunge, M. 1996. *Finding Philosophy in Social Science*, New Haven, CT: Yale University Press.

Cambridge Dictionary of Philosophy (2nd ed.). 1999. Cambridge, England: Cambridge University Press.

Cambridge International Dictionary of English (online version). (http://dictionary.cambridge.org/ dictionary/british/empirical; retrieved April 10, 2010).

Carroll, L. 1871. *Through the Looking-Glass, and What Alice Found There*, London: Macmillan.

Chalmers, A. F. 1999. *What Is This Thing Called Science?* (3rd ed.), Maidenhead, England: Open University Press.

Chattopadhyaya, D. P. 1991. *Induction, Probability, and Skepticism*, Albany, NY: State University of New York Press.

Cook, T. D., and Campbell, D. T. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Boston: Houghton Mifflin.

Copi, I. M., and Cohen, C. 1990. *Introduction to Logic* (8th ed.), New York: Macmillan.

Cronbach, L. J. 1975. "Beyond the Two Disciplines of Scientific Psychology," *American Psychologist* (30:2), pp. 116-127.

Cronbach, L. J. 1982. *Designing Evaluations of Educational and Social Programs*, San Francisco: Jossey-Bass.

Gefen, D., and Straub, D. W. 1997. "Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model," *MIS Quarterly* (21:4), pp. 389-400.

Gensler, H. 2001. *Introduction to Logic*, London: Routledge.

Ghoshal, S., and Moran, P. 1996. "Bad for Practice: A Critique of the Transaction Cost Theory," *Academy of Management Review* (21:1), pp. 13-47.

Gibbert, M., Ruigrok, W., and Wicki, B. 2008. "What Passes as a Rigorous Case Study?," *Strategic Management Journal* (29:13), pp. 1465-1474.

Gomm, R., Hammersley, M., and Foster, P. 2000. "Case Study and Generalization," in *Case Study: Key Issues, Key Texts*, R. Gomm, M. Hammersley and P. Foster (eds.), London: Sage Publications, pp. 98-115.

Harrison, D., and Easton, G. 2004. "Temporally Embedded Case Comparison in Industrial Marketing Research," in *Critical Realist Applications in Organisation and Management Studies*, S. Fleetwood and S. Ackroyd (eds.), London: Routledge, pp. 194-210.

Hofstede, G. 1980. *Culture's Consequences: International Differences in Work Related Values*, London: Sage Publications.

Hofstede, G. 1993. "Cultural Constraints in Management Theories," *Academy of Management Executive* (7:1), pp. 81-94.

Hume, D. 1975 [1748]. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, Oxford, England: Clarendon Press.

Hurley, P. 2003. *A Concise Introduction to Logic*, Belmont, CA: Wadsworth.

Kaplan, A. 1964. *The Conduct of Inquiry*, San Francisco: Chandler Publishing.

Klein, H. K., and Rowe, F. 2008. "Marshaling the Professional Experience of Doctoral Students: A Contribution to the Practical Relevance Debate," *MIS Quarterly* (32:4), pp. 675-686.

Knights, D. 1992. "Changing Space: The Disruptive Impact of a New Epistemological Location for the Study of Management," *Academy of Management Review* (17:3), pp. 514-536.

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

Lindgren, R., Henfridsson, O., and Schultze, U. 2004. "Design Principles for Competence Management Systems: A Synthesis of an Action Research Study," *MIS Quarterly* (28:3), pp. 435-472.

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

Majchrzak, A., Rice, R. E., Malhotra, A., King, N., and Ba, S. 2000. "Technology Adaptation: The Case of a Computer-Supported Interorganizational Virtual Team," *MIS Quarterly* (24:4), pp. 569-600.

Maxwell, J. A. 1992. "Understanding and Validity in Qualitative Research," *Harvard Educational Review* (62:3), pp. 279-300.

Merriam-Webster Dictionary (online version) (http://www.merriam-webster.com/dictionary/generalize, retrieved December 5, 2009, and http://www.merriam-webster.com/dictionary/empirical, retrieved April 10, 2010).

Nagel, E. 1979. *The Structure of Science: Problems in the Logic of Scientific Explanation*, Indianapolis, IN: Hackett Publishing.

Numagami, T. 1998. "The Infeasibility of Invariant Laws in Management Studies: A Reflective Dialogue in Defense of Case Studies," *Organization Science* (9:1), pp. 2-15.

Olsson, H. H., Ó Conchüir, E., Ågerfalk, P. J., and Fitzgerald, B. 2008. "Two-Stage Offshoring: An Investigation of the Irish Bridge," *MIS Quarterly* (32:2), pp. 257-279.

Oxford English Dictionary (2nd ed.). 1989. Oxford, England: Clarendon Press.

Ragu-Nathan, T. S., Tarafdar, M., and Rgau-Nathan, B. S. 2008. "The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation," *Information Systems Research* (19:4), pp. 417-433.

Reichenbach, H. 1938. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*, Chicago: University of Chicago Press.

Reichenbach, H. 1949. *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*, Berkeley, CA: University of California Press.

Rosenberg, A. 1993. "Hume and the Philosophy of Science," in *The Cambridge Companion to Hume,* D. Norton (ed.), New York: Cambridge University Press, pp. 64-89.

Rosenzweig, P. M. 1994. "When Can Management Science Research Be Generalized Internationally?," *Management Science* (40:1), pp. 28-39.

Rousseau, D. M., and Fried, Y. 2001. "Location, Location, Location: Contextualizing Organizational Research," *Journal of Organizational Behavior* (22:1), pp. 1-13.

Salmon, W. C. 1974. "The Pragmatic Justification of Induction," in *The Justification of Induction*, R. Swinburne (ed.), London: Oxford University Press, pp. 85-97.

Schofield, J. W. 1990. "Increasing the Generalizability of Qualitative Research," in *Qualitative Inquiry in Education: The Continuing Debate*, E. W. Eisner and A. Peshkin (eds.), New York: Teachers College Press, pp. 201-232.

Shadish, W. R., Cook, T. D., and Campbell, D. T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston: Houghton Mifflin.

Sharp, K. 1998. "The Case for Case Studies in Nursing Research: The Problem of Generalization," *Journal of Advanced Nursing* (27:4), pp. 785-789.

Stake, R. E., and Trumbull, D. J. 1982. "Naturalistic Generalizations," *Review Journal of Philosophy and Social Science* (7:1-2), pp. 1-12.

Strawson, P. F. 1952. *Introduction to Logical Theory*, London: Methuen & Co.

Tsoukas, H. 1989. "The Validity of Idiographic Research Explanations," *Academy of Management Review* (14:4), pp. 551-561.

Van Maanen, J. 1983. "The Fact of Fiction in Organizational Ethnography," in *Qualitative Methodology*, J. Van Maanen (ed.), Beverly Hills, CA: Sage Publications, pp. 37-55.

Venkatesh, V., and Ramesh, V. 2007. "Web and Wireless Site Usability: Understanding Differences and Modeling Use," *MIS Quarterly* (30:1), pp. 181-206.

Vickers, J. 2006. "The Problem of Induction," in *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu/entries/induction-problem; retrieved April 21, 2007).

Walton, J. 1992. "Making the Theoretical Case," in *What Is a Case? Exploring the Foundations of Social Inquiry*, H. S. Becker and C. Ragin (eds.), Cambridge, England: Cambridge University Press, pp. 121-137.

Yin, R. K. 2003. *Case Study Research: Design and Methods* (3rd ed.), Thousand Oaks, CA: Sage Publications.

About the Authors

Eric W. K. Tsang is the Dallas World Salute Distinguished Professor of Global Strategy in the School of Management, University of Texas at Dallas. He received his Ph.D. from the University of Cambridge. His research interests include organizational learning, strategic alliances, foreign direct investments, and philosophical analysis of research methodological issues. He is a consulting editor of *Management and Organization Review*, and is an editorial board member of the *Academy of Management Journal, Academy of Management Review, Journal of International Business Studies, Journal of Management Studies, Journal of World Business, Management International Review*, and *Organization Studies*.

John N. Williams is an associate professor in the School of Social Sciences, Singapore Management University. He received his Ph.D. from Hull University. His research interests include paradoxes, theory of knowledge, philosophy of religion, and applied ethics. His research has been published in leading journals of philosophy, such as *American Philosophical Quarterly, Analysis, Journal of Philosophical Research, Mind, Philosophical Studies,* and *Synthese.* He is a coeditor of *Moore's Paradox: New Essays on Belief, Rationality and the First Person*.

Appendix A

A Two-Stage Balance-of-Evidence Argument for Using Induction

We will *not* claim to have solved Hume's (1975) problem of induction. Instead we will use a two-stage argument that the balance of evidence is in favor of the verdict that there is some solution to the problem, even if we do not know what it is. In stage one, we will show that *there can be no evidence* that there is *no* solution because it is impossible to list in advance all putative solutions. Next we argue that at a minimum, there is one solution—and perhaps more—that has some degree of plausibility. We could establish this by giving a plausible defense of any of a number of solutions. A good place to pick one is Chattopadhyaya (1991), who gives an exhaustive treatment of the problem of induction. The ones we have picked are Strawson's (1952) and Reichenbach's (1938, 1949). If we have succeeded so far then we may go to the second stage to conclude that there is more reason to think that there is some successful solution out there than there is reason to think otherwise, even if we don't know what solution it is.

To begin the first stage, we have already noted above that Lee and Baskerville (2003) are committed to claiming that Hume's problem of induction is "irremediable" (p. 224). Yet they give us no reason for accepting this serious news. To show that Hume's problem of induction is irremediable—in other words, to give evidence that *there is no solution*—one would have to first *list all the candidate solutions*. Then one would have to demonstrate that each one fails, ideally winning general acceptance in every case. This is impossible because there is no way of knowing whether one's list is exhaustive. Who is to say that there is no infinity of distinct putative solutions? If the list is infinite then one could not finish it even if one were immortal. This would be like listing all the prime numbers. Even if the list is finite, demonstrating that one has excluded no distinct candidate solution is still impossible, because one cannot say in advance which nonidentical candidates may turn up to apply. In this respect, philosophers might say that the list is finite yet *indeterminate*. There may well be solutions that nobody has yet thought of. Some of these might be successful. Perhaps there are successful solutions that have been mistakenly written off as failures because nobody has yet given correct replies to compelling but ultimately misguided objections. But couldn't someone argue that Hume's problem has no solution by producing a long list of failed solutions? No, because that would be to use *induction*, by arguing from solutions we have observed to those we have not! Nothing lies in the scale on *this* side of the balance of evidence.

Note that we are not *expecting* anyone to *do* the impossible. All we are pointing out is that something *is* impossible, namely showing that there is no solution to the problem of induction. This is not simply because "There are no solutions that are successful solutions" is a universal negative claim of the form "no A are B." Some claims of this form are provable, for example "no integer larger than 10 is an integer less than 5" or "no citizen of Singapore is older than 100." But "there is no solution to the problem of induction" is impossible to prove because to prove it one would have to first *list all the candidate-solutions*. As we have argued, this is impossible.

Any opponent of our claim that there can be no evidence that there is no solution to the problem of induction must show not only why our reasons for this claim fail, but must also show that *there can be evidence* that there is no solution of the problem of induction. The prospects for any such response look dim.

Now consider what lies in the scale on the *other* side of the balance of evidence. Here the history of philosophy lies before us, with a plethora of attempts across a variety of traditions to grapple with Hume's problem. In the 14^{th} century, Duns Scotus effectively raised the same problem, to which Kant was a notable respondent. The first serious respondent to the problem raised in Hume's tradition was his contemporary, Thomas Reid. Modern respondents include Popper, Russell, Black, Nagel, Goodman, Strawson and Reichenbach, as well as postmodernists such as Derrida and Deleuze (see Chattopadhyaya 1991).

We think that Strawson's solution is plausible. He observes that to ask what it means to be justified in our beliefs is to ask what it means to base them on the evidence available.[10] What does *evidence* mean? There are two kinds of evidence, deductive and inductive. To say that a proposition is supported by deductive evidence is just to say that it can be deduced from propositions that we should accept. For example, if we accept the proposition that a given triangle is equiangular, then we are justified in believing the proposition that it is equilateral, because the latter proposition follows deductively from the former. Likewise, if we accept the vast body of inductive evidence available—that the law of gravity has always been observed to operate—then we are justified in believing that it will continue to operate. To ask whether we are justified in believing that the law of gravity will continue to operate is to ask whether we are justified in basing this belief on the evidence available, which, in this case, is inductive evidence. But basing our belief on evidence is simply what it *means* to be justified. To ask whether we should form beliefs on the basis of inductive evidence is tantamount to asking the trivial question, "Are we justified in being justified?" Trivially, the answer is yes. Once this is understood, Hume's problem is no problem at all.

An objection to this solution is that Strawson has conflated two senses of *justified*. One is the sense of justification for believing that induction is reliable. The other is the sense of *vindication*, in other words justification for using induction as a method of making as many true predictions as possible. Seen this way, the question whether we should form beliefs on the basis of inductive evidence amounts to the non-trivial question, "Are we justified in one sense in being justified in another?"

Our reply in defense of Strawson—one we think is original—is that this nontrivial question is either "If we have a good reason to believe that induction is reliable, then do we have a good reason for using induction as a method of making as many true predictions as possible?" or "If we have a good reason for using induction as a method of making as many true predictions as possible, then do we have a good reason to believe that induction is reliable?" The answer to the first is yes. The answer to the second is no, but as long as we have guidance for action, we need not lament the absence of logical proof. In other words, vindication without justification is better than neither.

In any case, Strawson goes on to vindicate induction. He observes rightly that no problem arises in the first place because induction is impossible to vindicate only if we impose deductive standards on it. But induction by definition is not deduction, so we should not expect induction to meet deductive standards. Instead we should judge induction by its own standards. What justifies the practice of deduction is that deduction is *truth preserving*—if you put truth in, you are sure of getting truth out. What justifies the practice of induction is that induction is *ampliative* in the sense that it "can amplify and generalize our experience, broaden and deepen our empirical knowledge" (Vickers 2006); because we venture into the territory of the unobserved, we may hope to add to our knowledge. These are complementary virtues, and we should not expect each type of inference to embody both. We can't have our cake and eat it.

A pragmatist might add that we need not vindicate induction to justify its use. Instead of trying to show that induction is a method of making as many true predictions as possible, we should look to the expected benefit of choosing to use it. Such a pragmatist might endorse Reichenbach's solution, described and defended by Salmon (1974). Faced with the need to predict the future, a scientist may use induction or some rival method such as crystal-ball gazing. Which should she decide to use? Either nature is uniform (in the sense that laws of nature, once true, are always true) or it is not. If not, then no method of predicting the future will work, including induction and its rivals. This is because if the rival method worked—in other words, if it consistently yielded true predictions—then its success would constitute a uniformity that could be exploited by the inductive method. For example, we could inductively infer the future success of the crystal-ball gazer from her past successes. On the other hand, if nature is uniform, then induction will definitely work, but rival methods, such as crystal-ball gazing, might or might not work. "We have, therefore, everything to gain and nothing to lose by using induction. If induction fails, no other method could possibly succeed" (Salmon 1974, p. 86). Faced with a choice between the chance of guidance and the certainty of no guidance at all, the former choice is the best. By analogy, before venturing into yet unobserved territory, it is more sensible to use a map that may or may not be reliable than to throw it away.[11]

---

[10]Strawson uses the term *reasonable* as a synonym of *justified*. Nothing turns upon this.

[11]This of course is not to construe the choice as between a map which is known to be unreliable and no map at all. In that scenario, no map might be the better choice.

As Chattopadhyaya shows, these are not the only voices in the debate. For example, it might be argued that Strawson's view is consistent with that of Pierce, James, Toulmin, or even Hume himself. Moreover it would be rare to find a solution to a deep philosophical problem that cannot be criticized from a variety of perspectives. Such is the nature of philosophy. Perhaps we are wrong to conclude that Strawson or Reichenbach has an adequate solution. But in that case, we should think of them as illustrations of how two putative solutions—among many others—might be defended. They provide at least *some* evidence that there is some adequate solution out there.

We have shown that it is impossible to demonstrate that there is *no* solution—because there is no way of identifying all candidate solutions in advance. We have also shown that there is a minimum of a single solution that has some degree or other of plausibility. We may therefore conclude that there is more reason to think that there is some successful solution out there than there is reason to think otherwise, even if we don't know what solution it is.

Appendix B

A Summary of "Generalizing Generalizability in Information Systems Research" ■

Lee and Baskerville (2003) define generalizability and argue that generalizability need not have a quantitative or statistical dimension. However, many IS researchers have just restricted themselves to statistical, sampling-based generalizability and inappropriately applied this notion of generalizability to nonstatistical, nonsampling forms of research. The purpose of the article is "to clarify the concept of generalizability by critically examining its nature, illustrating its use and misuse, and offering a framework for classifying its different forms" (p. 221).

In section 2, Lee and Baskerville invoke Yin's (2003) distinction between Level 1 and Level 2 inference in his seminal work *Case Study Research: Design and Methods*, and endorse his argument that statistical generalizability belongs to Level 1 inference whereas inference from case study findings to theory is Level 2 inference. Thus statistical generalizability is not an appropriate measure of the quality of case studies. In their Table 1 (p. 223), they list examples of published studies that inappropriately apply the conception of statistical generalizability to case research.

In section 3, they examine inductive reasoning, of which statistical generalizing is one form, and focus on Hume's (1975) problem of induction. They commit themselves to the claim that the problem is "irremediable" (p. 224), and thus that statistical generalizability is not justified. That is, an increase in sample size will not improve the generalizability of a sample to its population. Yet they claim that Hume's problem of induction imposes "no prohibition of the conclusion that an increase in sample size leads to an increase in the generalizability of one sample to other samples that the same sampling procedure would produce" (p. 228).

They then examine different conceptions of generalization in positivism and interpretivism in section 4. A key feature of positivism is its emphasis on generalization with an aim to discover invariable universal laws, and statistical generalization falls within the domain of positivism. In contrast, interpretivism places no particular emphasis on generalization but would not prohibit researchers from extending their theory from the setting where it was developed to other settings.

They classify generalization into four types in section 5. They first make a distinction between empirical and theoretical statements, both of which are made use of in positivist or interpretivist research. They then classify generalization as empirical statement to empirical statement (EE), empirical statement to theory (ET), theory to empirical statement (TE), and theory to theory (TT), and elaborate on each. In particular, they discuss statistical generalization as Type EE. They also illustrate their classification with Gefen and Straub's (1997) article, "Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model," in section 6.

In the conclusion they reiterate their crucial findings:

(1) In neither case studies nor statistical studies, would it be appropriate to criticize a theory for a lack of generalizability to other settings.

(2) Neither an increase in the sample size in a statistical study nor an increase in the number of sites in a case study would increase the generalizability of a theory.