MIS Quarterly

ISSUES AND OPINIONS

CONCEPTUALIZING GENERALIZABILITY: NEW CONTRIBUTIONS AND A REPLY

Allen S. Lee
Virginia Commonwealth University, 301 West Main Street, Richmond, VA 23284-4000 U.S.A. {allenlee@alum.mit.edu}

Richard L. Baskerville
Georgia State University, POB 4015, Atlanta, GA 30302-4015 U.S.A. {baskerville@acm.org}

Appendix A

How to Generalize a Theory to a New Setting

This appendix provides additional details of the illustration that the essay has already provided, in part, for how to generalize a theory to a new setting. The theory is one originally developed and empirically tested by Markus (1983). The illustration involves the generalizing of Markus' theory from the organization Markus called GTC to a new organization that we call QED. The additional details build on Tables 1, 2, and 3, which identified the six variables and eight statements making up Markus' theory. Table 3 provided the details of generalizing Markus' theory for just two of the six variables and one of the eight propositions. In the additional Tables A1 and A2, we now present the details for generalizing which involves all six variables.

Table A1. The Empirical Values Taken by the Theory in the Original Setting (GTC) from Which the Theory Is Generalized

OPD	organizational power distribution, as instantiated in the situation where GTC's divisional accountants have power to control access to financial data, but GTC's corporate accountants lack power to control access to financial data
ICB	intention to change the balance of power, as instantiated in the situation where the intention of Howard, head of GTC's corporate controller's office, is to change the balance of power from GTC's Chemical Divisions (where his rival, Spade, is GTC's Chemical Company Controller) to GTC's corporate management, and where GTC's corporate accountants intend "from the very beginning that FIS be used for managerial accounting, not just ... financial accounting" (p. 439)
PSSD	power shift implied in the system design, as instantiated in the situation where the design of FIS supports giving power to GTC's corporate accountants and removing power from GTC's divisional accountants
R	resistance, as instantiated in the situation where GTC's divisional accountants resist FIS
PT	political tactics, as instantiated in the situation where GTC's corporate accountants have intentions of using FIS for managerial accounting but do not "immediately reveal these intentions to the divisions" at GTC; where some of GTC's divisional accountants continue using their old accounting methods ("thick manual ledger books") in parallel with FIS; where some of GTC's divisional accountants engage in "data fudging" when using FIS; and where GTC's corporate accounting makes adoption of FIS voluntary but requires non-adopters to report the same information manually.
PSR	power shift realized in the organization, as instantiated in the situation where there is a shift in power from GTC's divisional accounting to GTC's corporate accounting (i.e., in the situation where Spade retires and his old position of GTC's Chemical Company Controller is "moved under the direct line control of [GTC's] corporate accounting and then eliminated," hence realizing a shift in power from GTC's divisional accounting to GTC's corporate accounting).

Now, how would a researcher, working with a practitioner at a different organization, QED, generalize Markus' theory from GTC to QED? The cross-out font and the underlining in Table A2 indicate the changes in values taken by the six variables in this generalization.

Table A2. The Empirical Values Taken by the Theory in the New Setting (QED) to Which the Theory Is Conservatively Generalized

OPD	organizational power distribution, as instantiated in the situation where ~~GTC's~~ QED's divisional accountants have power to control access to ~~financial~~ valued data, but ~~GTC's~~ QED's corporate accountants lack power to control access to ~~financial~~ valued data
ICB	intention to change the balance of power, as instantiated in the situation where the intention of ~~Howard~~ Reginald, head of ~~GTC's~~ QED's corporate controller's office, is to change the balance of power from ~~GTC's~~ QED's ~~Chemical~~ Manufacturing Divisions (where his rival, ~~Spade~~ Nigel, is ~~GTC's~~ QED's ~~Chemical~~ Manufacturing Company Controller) to ~~GTC's~~ QED's corporate management, and where ~~GTC's~~ QED's corporate accountants intend "from the very beginning that ~~FIS~~ AIS [QED's new Accounting Information System] be used for managerial accounting, not just ... financial accounting" (p. 439)
PSSD	power shift implied in the system design, as instantiated in the situation where the design of ~~FIS~~ AIS supports giving power to ~~GTC's~~ QED's corporate accountants and removing power from ~~GTC's~~ QED's divisional accountants
R	resistance, as instantiated in the situation where ~~GTC's~~ QED's divisional accountants resist ~~FIS~~ AIS
PT	political tactics, as instantiated in the situation where ~~GTC's~~ QED's corporate accountants have intentions of using ~~FIS~~ AIS for managerial accounting but do not "immediately reveal these intentions to the divisions" at ~~GTC~~ QED; where some of ~~GTC's~~ QED's divisional accountants continue using their old accounting methods ("thick manual ledger books") in parallel with ~~FIS~~ AIS; where some of ~~GTC's~~ QED's divisional accountants engage in "data fudging" when using ~~FIS~~ AIS; and where ~~GTC's~~ QED's corporate accounting makes adoption of ~~FIS~~ AIS voluntary but requires non-adopters to report the same information manually.
PSR	power shift realized in the organization, as instantiated in the situation where there is a shift in power from ~~GTC's~~ QED's divisional accounting to ~~GTC's~~ QED's corporate accounting (i.e., in the situation where ~~Spade~~ Nigel retires and his old position of ~~GTC's~~ QED's ~~Chemical~~ Manufacturing Company Controller is "moved under the direct line control of [~~GTC's~~ QED's] corporate accounting and then eliminated," hence realizing a shift in power from ~~GTC's~~ QED's divisional accounting to ~~GTC's~~ QED's corporate accounting).

As mentioned previously, the ethical researcher must make explicit to the practitioner who seeks to apply the theory that there are risks associated with (1) generalizing "GTC divisional accountants" to "QED divisional accountants," (2) generalizing "financial data" to "valued data," (3) generalizing "Howard, head of GTC's corporate controller's office" to "Reginald, head of QED's corporate controller's office," (4) generalizing "GTC Chemical Divisions" to "QED Manufacturing Divisions," and (5) generalizing "FIS," the new accounting system at GTC, to "AIS," the new accounting system at QED.

Suppose, however, even further generalizing is required or desired. As mentioned in the essay, for example, suppose that the new setting QED has no "divisional accountants" to which the actions associated with GTC's "divisional accountants" can be generalized. Table A3 illustrates the form that the additional generalizing can entail for all six variables, where the cross-out font, the cross-out font with wavy underlining, and the plain underlining indicate the change from conservative generalizing to liberal generalizing. A reader, by following the progression of changes in fonts, can see how the generalizing becomes more liberal.

Table A3. The Empirical Values Taken by the Theory in the New Setting to Which the Theory Is Liberally Generalized

OPD	organizational power distribution, as instantiated in the situation where ~~GTC's~~ ~~QED's divisional accountants~~ managers in one division of QED have power [henceforth, the QED division initially with power] have power to control access to ~~financial~~ valued data, but ~~GTC's~~ ~~QED's corporate accountants~~ managers in another division of QED [henceforth, the QED division initially without power] lack power to control access to ~~financial~~ valued data
ICB	intention to change the balance of power, as instantiated in the situation where the intention of ~~Howard~~ Reginald, head of ~~GTC's~~ QED's corporate controller's office, is to change the balance of power from ~~GTC's~~ QED's ~~Chemical~~ Manufacturing Divisions (where his rival, ~~Spade~~ Nigel, is ~~GTC's~~ QED's ~~Chemical~~ Manufacturing Company Controller) to ~~GTC's~~ QED's corporate management, and where ~~GTC's~~ QED's corporate accountants intend "from the very beginning that ~~FIS~~ AIS [QED's new Accounting Information System] be used for managerial accounting, not just ... financial accounting" (p. 439)
PSSD	power shift implied in the system design, as instantiated in the situation where the design of ~~FIS~~ AIS supports giving power to ~~GTC's~~ ~~QED's corporate accountants~~ managers in the QED division initially without power and removing power from ~~GTC's~~ QED's divisional accountants managers in the QED division initially with power
R	resistance, as instantiated in the situation where ~~GTC's~~ ~~QED's divisional accountants~~ managers in the QED division initially with power resist ~~FIS~~ AIS
PT	political tactics, as instantiated in the situation where ~~GTC's~~ ~~QED's corporate accountants~~ the managers in the QED division initially without power have intentions of using ~~FIS~~ AIS for managerial accounting but do not "immediately reveal these intentions to the divisions" at ~~GTC~~ QED; where some of ~~GTC's~~ ~~QED's divisional accountants~~ the managers in the QED division initially with power continue using their old accounting methods ("thick manual ledger books") in parallel with ~~FIS~~ AIS; where some of ~~GTC's~~ ~~QED's divisional accountants~~ the managers in the QED division initially with power engage in "data fudging" when using ~~FIS~~ AIS; and where ~~GTC's~~ ~~QED's corporate accounting~~ the QED division initially without power makes adoption of ~~FIS~~ AIS voluntary but requires non-adopters to report the same information manually.
PSR	power shift realized in the organization, as instantiated in the situation where there is a shift in power from ~~GTC's~~ ~~QED's divisional accounting~~ the QED division initially with power to ~~GTC's~~ ~~QED's corporate accounting~~ the QED division initially without power (i.e., in the situation where ~~Spade~~ Nigel retires and his old position of ~~GTC's~~ QED's ~~Chemical~~ Company Controller is "moved under the direct line control of ~~[GTC's~~ ~~QED's] corporate accounting~~ the QED division initially without power and then eliminated," hence realizing a shift in power from ~~GTC's~~ ~~QED's divisional accounting~~ the QED division initially with power to ~~GTC's~~ ~~QED's corporate accounting~~ the QED division initially without power).

In Table A3, every leap in generalizing (from the cross-out font, to the cross-out font with wavy underlining, to the plain underlining) diminishes the similarity that the uniformity of nature assumption requires to exist between the setting being generalized from (GTC) and the setting being generalized to (QED). This, in turn, diminishes the confidence that one may have in the generalizing that one is performing, whether this is seen from T&W's perspective or L&B's perspective.

Also note that the construction of tables similar to Table 3, indicating how the remaining theoretical statements change when generalized, would employ the same reasoning as Tables A2 and A3.

Appendix B

Compatibility Between Sampling and Hume's Problem of Induction

The compatibility between accepting Hume's problem of induction and accepting the statistical concept of sampling is such a critical (and, as we have learned from discussions with our colleagues, misunderstood) point that it calls for a mathematical explanation, which follows.

Consider the case of interval estimation for the population proportion p, whose true value is unknown but is estimated with a sample proportion $\bar{p}$ (see Neter et al. 1988, pp. 279-286, 367-382). A confidence interval can be denoted as $L \leq p \leq U$, where L is the confidence interval's lower bound $\left(L = \bar{p} - z_{(1-\alpha/2)}\sqrt{\bar{p}(1-\bar{p})/(n-1)}\right)$, U is its upper bound $\left(U = \bar{p} + z_{(1-\alpha/2)}\sqrt{\bar{p}(1-\bar{p})/(n-1)}\right)$, n is the sample size, α is the level of statistical significance, and $z_{(1-\alpha/2)}$ is the value of the standard normal random variable Z such that probability $P(Z \leq z) = 1 - \alpha/2$.

A given confidence interval is correct if the population proportion p falls between the interval's bounds L and U. A researcher who will take a sample and then construct a confidence interval does not know whether it will be correct, but the researcher can compute and assign the probability that it will be correct as follows:

$$P(L \le p \le U) = P\left(\bar{p} - z_{(1-\alpha/2)} \sqrt{\frac{\bar{p}(1-\bar{p})}{n-1}} \le p \le \bar{p} + z_{(1-\alpha/2)} \sqrt{\frac{\bar{p}(1-\bar{p})}{n-1}} \right) = 1 - \alpha$$

Consider the following situation: "many random samples of size n [are taken] from a population" (Neter et al. 1988, p. 285); the sample size n is a known number and it is large; and the desired maximum error or difference, $z_{(1-\alpha/2)} \sqrt{\bar{p}(1-\bar{p})/(n-1)}$, between a sample proportion and the population proportion is specified at one or another value. For the "many random samples" just mentioned, suppose that 95 percent of them result in confidence intervals that are correct—that is, 95 percent of them include the actual but unknown value of the population proportion p between their lower and upper bounds. In other words, $P(L \le p \le U) = 1 - \alpha = .95$.

Next, consider the same situation, but with just one change. The sample size of each of the "many random samples" just mentioned is no longer n, but a particular number greater than n. The intuitive result is that, with the increase in sample size, there will also be an increase in the percentage of confidence intervals that are correct. The intuitive result is confirmed by the mathematical rationale that, where the desired maximum error or difference $z_{(1-\alpha/2)} \sqrt{\bar{p}(1-\bar{p})/(n-1)}$ is not changed, an increase in n brings about a decrease in $\sqrt{\bar{p}(1-\bar{p})/(n-1)}$, which in turn requires $z_{(1-\alpha/2)}$ to increase, which happens through a decrease in α. As a result, $P(L \le p \le U)$, which is equal to $(1 - \alpha)$, becomes greater than .95. Such is the reasoning behind the practical consequence that, for an opinion poll where a desired maximum error is specified (e.g., plus or minus 3 percentage points), an increase in the sample size would be beneficial. This reasoning supports (and does not deny, as T&W claim) the conclusion that (in T&W's phrasing) "an increase in sample size increases the probability that the population has the characteristic found in the sample" (p. 730) or (in our preferred phrasing) an increase in sample size increases the probability that the confidence interval will include the population characteristic.

Where we define *reliability* as "the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials,"[1] a larger sample size n will result in greater reliability not only in interval estimation (i.e., a larger percentage of the interval estimates will include p and therefore be correct), but also in all statistical hypothesis testing which can be conducted with, or is functionally equivalent to, interval estimation.

This reasoning behind reliability supports the following statements (Lee and Baskerville 2003, p. 227):

> For a sample (one of the 100 [samples] in our example) that leads to the correct decision of not rejecting the null hypothesis, an increase in sample size would result in a larger number of other samples [those also leading to the correct decision of not rejecting the null hypothesis] to which this sample could itself be generalized. Therefore, a larger sample size does increase generalizability, but it is the generalizability of a sample to other samples, not to the population.

Of course, in L&B's example, a researcher does not know which ones of the other 100 samples are those leading to the correct decision of not rejecting the null hypothesis; the researcher only knows that, among the 100 samples, an increase in the size of each sample will increase the number or proportion of samples that will lead to the same decision as the researcher's sample and that, therefore, are those to which the researcher's sample can be generalized.

Note that the claim that a larger sample size increases the number or percentage of other samples also leading to the same result (i.e., the same result of including the population parameter between the bounds of the confidence interval or the same result of making the correct decision not to reject the null hypothesis) does not involve induction and therefore is not affected or prohibited in any way by Hume's problem of induction. Rather than induction, this involves reliability. A larger sample size in statistical inference is, therefore, beneficial insofar as it enhances reliability.

[1] http://www.merriam-webster.com/dictionary/reliability

Appendix C

Intractability of the Problem of Induction

Hume's problem of induction has survived hundreds of years of challenges in the philosophy of science. The arguments that have been marshaled against Hume's problem have failed to gain widespread acceptance among our colleagues in the philosophy of science. For example:

- From the *Encyclopædia Britannica* (2010, p. 1): "Philosophers have responded to the problem of induction in a variety of ways, though none has gained wide acceptance."

- From the *Oxford Companion to Philosophy* (Honderich 2005, p. 432): "Most philosophers hold that there is a problem about induction: its classic statement is found in Hume's *Enquiry Concerning Human Understanding* [and]….None of these supposed justifications [of induction] is universally accepted."

- Rosenberg (2005, pp. 141-142) states: "'Confirmation Theory,' as this part of the philosophy of science [logical positivism] has come to be called … has left as yet unsolved Hume's problem of induction."

- From *The Fontana Dictionary of Modern Thought* (Bullock and Stallybrass 1977, p. 306): "The justification of induction has been a persistent problem."

- Salmon, who published a paper in 1974 that T&W expressly cite in defense of induction, subsequently wrote in *The Cambridge Dictionary of Philosophy* (1999, p. 746): "Philosophers have responded to the problem of induction in many different ways.…None of the many suggestions is widely accepted as correct."

While not alone, T&W's opposition is not a widely held position. They ask, "Have Lee and Baskerville escaped Hume's problem of induction?" (p. 732). Hopefully not. L&B accepted Hume's problem of induction, exactly as Donald T. Campbell did. Given the centuries of debate on the matter, Hume's problem cannot be dismissed lightly and, indeed, the consensus is that it cannot be dismissed at all. Furthermore, it appears to us that the wide appreciation in the philosophy of science of Hume's problem reaches even more deeply than L&B earlier supposed: one must also contend with Goodman's (1955) "new riddle of induction." Even if Hume's problem of induction were to be solved someday, Goodman's riddle of induction would still be waiting to be solved. Hume's problem helps us to understand science at its edges and, rather than isolating science from non-science, it raises our appreciation of the value of diverse forms of science.

Nevertheless, T&W sweep past Hume's problem of induction with two arguments in their appendix. They do this, first, with a fundamentally inductive argument that operates as follows: All past observations of attempted solutions to Hume's truism have been observations of attempts that failed; therefore, the next observation will be an observation of failure; however, there is no uniformity of nature; therefore, because all past observations were observations of failures, the next observation will be an observation of success (or the one after that, or after that, etc.). This is similar to the "scientific" logic that promises a future solution to the disposal of nuclear waste: The future will solve this problem. Even Strawson (1952), some of whose reasoning T&W cite, challenges such logic: "But this is itself a higher-order induction: where irregularity is the rule, expect further irregularities. Learning not to count on things is as much learning an inductive lesson as learning what things to count on" (p. 262).

In a second argument, T&W recite an argument (based on Strawson) that draws forward the notion that evidence can mount probabilistically to provide justification for acting on beliefs about the future. However, belief is not the same as proof. Likewise, many observations may provide support, but support is also not the same as proof. Furthermore, there is still the current consensus in the philosophy of science that Hume's problem of induction remains unsolved—a consensus that does not attest to the viability of any proposed solution based on Strawson's over-half-century-old argument.

Appendix D

Defining Generalizability

T&W choose to define generalizability in a sense that can be situated narrowly within positivist science precepts. Since no one owns the language, this is perfectly suitable for their argument. This choice suits their purpose and seems appropriate to their aims.

In contrast, in L&B, we were aiming to develop a generalized form of the concept of generalizability—one elastic enough to stretch across diverse forms of scholarly research, including positivist science. For this purpose it is not surprising that L&B considered the most general definitions of the term, including those in common, everyday English usage and research usage.

Contrary to T&W's assertion, L&B did not adopt, but reported the *Oxford English Dictionary*'s (2003) definition of generalizability, just as L&B reported Yin's definition of it. L&B developed and contributed their own definitions of generalizing and generalizability (consisting of four types), where L&B's choice of definitions suited their purpose and was appropriate to their aims. In particular, L&B even defined one of their four types of generalizability as being the opposite of the *OED* definition:

> Type TE generalizability, which involves generalizing a theory confirmed in one setting to descriptions of other settings—refers to reasoning from theoretical statements to empirical statements, which is actually deduction, not induction. Indeed, such reasoning is the opposite of the OED definition of generalize, which is 'to form general notions by abstraction from particular instances' (p. 241).

In response to T&W, we nonetheless reaffirm that the *OED* definition describes generalization in one of the senses used by some influential researchers in the natural and social sciences. For instance, Shadish et al. (2002) include in their definition of generalizing: "To infer from many particulars, … To draw inferences or a general conclusion from. … To make generally or universally applicable. … To popularize" (p. 341).

Appendix E

Stereotyping, Induction, and the Statistical Syllogism

Many stereotypes arise from scientific findings about group differences because logical and statistical reasoning can mask motivated biases (Schaller 1992). The logic of stereotyping is present in what T&W identify, from earlier literature, as the "statistical syllogism":

> 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."

We point out that the logic of stereotyping is present in the following application of the statistical syllogism, where N = 75, F = (a woman working at Golden Triangle Corporation), G = (a secretary), and X = (Carol).

> P1 75% of the women working at Golden Triangle Corporation are secretaries.
> P2 Carol is a woman working at Golden Triangle Corporation.
> C Carol is a secretary.

Carol, however, is actually the CEO at Golden Triangle Corporation. The immediately preceding logic renders Carol a victim of gender stereotyping. And to the extent that "Carol is a secretary" does not necessarily follow from the two premises (P1 and P2), one may argue that the logic leading to this "conclusion" (C) is not well described as the logic of the syllogism. Information about stereotyped groups plays a critical role in inductive judgment because stereotyped features interact with the observer (Yamauchi 2005).

Of course, the logic of the syllogism does apply in statistics. Lee and Hubona (2009, p. 249) offer an example of this. The major premise (P1) is "if H_0 is true, then p-value > α," the minor premise (P2) is "p-value < α," and the conclusion (C) is "therefore, reject H_0 as true." Here, H_0 refers to the null hypothesis and α refers to the critical p-value or the desired level of statistical significance. This clarifies how the logic of statistical inference is deductive, not inductive.

References

Bullock, A., and Stallybrass, O. (eds.). 1977. *The Fontana Dictionary of Modern Thought*, London: Fontana.

Encyclopedia Britannica. 2010. "Problem of Induction," *Encyclopædia Britannica Online* (http://www.britannica.com/EBchecked/topic/1311323/problem-of-induction).

Goodman, N. 1955. *Fact, Fiction, & Forecast*, Cambridge, MA: Harvard University Press.

Honderich, T. (ed.). 2005. *The Oxford Companion to Philosophy*, Oxford, UK: Oxford University Press.

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

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

Lee, A., and Hubona, G. 2009. "A Scientific Basis for Rigor in Information Systems Research," *MIS Quarterly* (33:2), pp. 237-202.

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

Markus, M. L. 1983. "Power, Politics, and MIS Implementation," *Communications of the ACM* (26:6), pp. 430-444.

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

Rosenberg, A. 2005. *Philosophy of Science: A Contemporary Introduction*, New York: Routeledge.

Salmon, W. C. 1999. "Problem of Induction," in *The Cambridge Dictionary of Philosophy,* R. Audi (ed.), New York: Cambridge University Press, pp. 745-746.

Schaller, M. 1992. "In-Group Favoritism and Statistical Reasoning in Social Inference: Implications for Formation and Maintenance of Group Stereotypes," *Journal of Personality & Social Psychology* (63:1), pp. 61-74.

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

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

Tsang, E. W. K., and Williams, J. N. 2012. "Generalization and Hume's Problem of Induction: Misconceptions and Clarifications," *MIS Quarterly* (36:4), pp. 729-748.

Yamauchi, T. 2005. "Labeling Bias and Categorical Induction: Generative Aspects of Category Information," *Journal of Experimental Psychology: Learning, Memory & Cognition* (31:3), pp. 538-553.