Harvard Business School Research Colloquium

The Information Systems Research Challenge: Experimental Research Methods

Volume 2

Izak Benbasat,
editor

James I. Cash, Jr.
Jay F. Nunamaker, Jr.
series editors

Harvard Business School, Boston, Massachusetts

Chapter 10

Addressing the "Third Criterion" in Experimentalist Research: Towards a Balance of Manipulative and Analytic Control

T. S. Palys
School of Criminology
Simon Fraser University

Introduction

When social scientists perform research, they engage in a process that links theory and data in a setting chosen for its heuristic capability. Research offers a means of getting somewhere, the "somewhere" of interest being defined by the investigator in the context of general scientific objectives. Regardless whether we seek longer-term "Truths" or principles of short term, contemporary interest, our identity as social scientists is a reaffirmation of our belief that the interplay of theory and data is best served when research sites are strategically chosen and procedures are both systematic and planned.

Traditional social science lore has it that the heuristic capability of research settings is maximized when we exert control. More specifically, a variety of classic works suggests that it is through *manipulative* control that our objectives are most effectively addressed (e.g., see Boring, 1969; Campbell and Stanley, 1963; Kerlinger, 1973). Not surprisingly, hierarchies of research methods created in textbooks and the like are usually expressed in terms of the extent to which manipulative control can easily be exerted; laboratory experiments are "best," followed by field experiments, natural experiments, field studies, and case studies. Although field and case studies are often given at least token points for their "richness" and "realism," our affection is short-lived when we get to the *real* business of science, which is said to require allegiance to the desiderata of manipulative control, random assignment, aggregate data, and experimental and control groups.

The present paper argues that the hierarchy depicted above is problematic, largely because it treats manipulative control as an objective in itself. Evaluating experiments relative to other methods in terms of how well they employ manipulative control is little more than tautology; any method is "best" when evaluated according to its own criteria. Independent criteria are required. Specifically, methods should be evaluated in terms of their heuristic value and the range and clarity of inference they allow, since it is inference rather than manipulative control *per se* that we are attempting to achieve. From this perspective, manipulative control and its manifestations (e.g., random assignment) are seen in somewhat the same vein as computers, clipboards, and checklists--useful, to be sure, but mere *means* towards an end rather than ends in themselves. Although "control" is a useful concept when considering the merits of rival plausible explanations, there are different types of control--*manipulative* and *analytic* are discussed here--that can serve these ends. Our task is to unearth the strengths and weaknesses inherent in a particular research site or procedure in terms of its inferential capabilities, and that is what I attempt to do here.

The Experimental Vision

Historically in the social sciences, and nowadays in the Information Systems (IS) realm and beyond, we perceive experimentation, and laboratory experimentation in particular, as the route to the generation of reliable, dependable, and perhaps practical, knowledge. Although research embodying experimentalist principles was being conducted long before our current canons of research were formulated (e.g., see Cook and Campbell, 1979; Mason, 1988), our major contemporary debt in this realm is to John Stuart Mill, who gave us a reasonable and workable set of criteria to consider when evaluating whether two or more variables of interest are in some sense causally related. These include:

1. *Temporal Precedence,* i.e., we believe that causes come *before* effects, and hence must demonstrate that the putative cause did indeed occur *prior* to any changes or differences allegedly produced by it;

2. *Association* or *Relationship*, i.e., if the alleged cause does indeed act to produce a given effect, we must show that changes in a particular independent variable covary in some reliable way with changes that are observed in a dependent variable; and

3. *Elimination of Rival Plausible Explanations*, i.e., we are obliged to demonstrate that it is the putative cause *per se* that is responsible for changes in the dependent measure, rather than related variables, nuisance variables, artifacts, or any of myriad other potential causal agents that might have been present.

The contemporary laboratory experiment, which emerged directly from Mill's exposition, does a splendid job of addressing these criteria. A brief review of how it does so is useful here.

Experimentation begins when we recognize or create a situation that includes the phenomenon of interest to us and embodies parameters suggested by theory, or includes the intervention whose impact we wish to address. The situation or intervention need not exist in reality, implying that the lab affords us the ability to deal not only with realities that exist, but also to consider *possible* realities that *might* exist if our particular technology or organizational structures were implemented, or if the world were arranged in the manner imagined by us or our theories. Because the laboratory experiment is conducted on our turf, as we arrange it, and when we want it to happen, we seize the luxury of causing the cause to occur at our convenience, enabling us to be there with our dependent measures in hand, waiting to catch the anticipated effects when they occur.

The dicta of experimental practice help to ensure that our labors will be fruitful in meeting Mill's criteria. A simple before-after design will show us that the dependent measure did not change until after the alleged cause was imposed; random assignment coupled with the existence of a treatment and a control group in a post-test only design leaves us confident that any differences that might exist *after* the experimental intervention were not there before. Either way, the temporal precedence criterion will have been met. In the process, we also will have demonstrated that a relationship exists between the presence/absence of the independent variable and the presence/absence of the dependent measure; whenever we press the "cause" button, an effect emerges at the other end. Karl Pearson, Sir Ronald Fisher, and their intellectual descendants have armed us with assorted statistical techniques to facilitate our decisions concerning whether the magnitude of change or difference we observe is something to pursue further, or merely within the realm of variation that might be expected on the basis of chance processes alone.

But showing temporal precedence and the existence of a relationship is not enough. It only brings us to the tough part, the third criterion, which challenges us to show that the variation or change we observed was specifically due to the independent variable *per se*, as opposed to any of a multitude of other related and/or spurious causal agents that might have been present. This is important. Because of the investment we and/or sponsors have in the technology under consideration, or because of the import we have speculatively ascribed to the theoretical constructs embodied in our variables, we want to be as squeaky-clean as possible in identifying the locus of change. In order to fulfill this obligation, we turn to that panacea called control.

Since that first experiment of Aristotle's (Mason, 1988), and particularly during the last century, social scientists have learned much about control. The laboratory allows us to create a closed system under our surveillance, minimizing from the start the number and sorts of variables that might otherwise impinge upon events. We create control/comparison groups to serve as a baseline for change and develop measures that are reliable, valid, and sensitive enough to detect change, if it should occur. By randomly assigning participants to groups, we ensure that the average units of each group are theoretically equivalent in all respects.

The experimenter assumes control of defining the situation; the participants, who are generally prepared to suspend disbelief and do their best, are asked to listen attentively to our instructions. To ensure that any effects we might observe are not the result of intentional or unintentional variations in procedure, we seek consistency and precision in the way the procedures are administered. Consequently, we might tape-record our instructions, show them on a monitor, and/or have a hard copy available to which participants can refer. As experimenter expectancies can exert an influence, the experimenter might be kept blind about group memberships and/or the nature of hypotheses being tested. Although random assignment will have equated treatment and control groups on all pre-experimental variables, we will take steps to ensure that we create no new differences incidental to the treatment being assessed. We may not tell participants the definitive characteristics of their group, nor of others. Finally, after defining the situation, indicating the constraints under which they must operate, and providing what we perceive to be appropriate means through which they might respond, we let the participants behave.

In the end, we will have created two groups (in the simplest case), equivalent in all respects but one: the presence or absence of the independent variable whose impact we wish to assess. A sort of *ceteris paribus* purity is created, and the test is set. If the difference or change is statistically significant under such conditions, we label the independent variable efficacious, since our high internal validity identifies that variable as the only possible agent of change. If the result is nonsignificant, our confidence in the alleged causal variable's efficacy is undermined.

Our attention turns next to the *ceteris* we have held *paribus*. Could it be that the observed relationship holds only under the conditions in which, and for the participants with whom, the original experiment was executed? We begin to focus on possible variations in these parameters, and on what these might imply in the way of either caveats or bolder affirmations. Does the theoretical relationship hold across other persons, settings, and times? These are, of course, questions of external validity, which, although they might be argued on a purely rational basis, ultimately come down to empirical demonstrations in subsequent research. Ideally, we hope that these demonstrations will be done with the same degree of precision and control as our initial study.

I doubt sincerely that any of the foregoing is news to its readers. It is intended as a summary of the logic that underlies experimentation and, I hope, will serve as evidence that there is some overlap in our understanding of what constitutes "good" experimentation. If I haven't yet said anything you consider contentious, we are in good shape to begin questioning some of these chestnuts of experimentation.

The Alchemy of Experimentation

When first we set out with the objectives of developing theory and/or testing the impact of interventions in the IS realm, we are immediately beset by decisions concerning the general approach to adopt. Mason (1988) is not alone, I believe, in suggesting that the general trade-off we must consider is the one of whether we will emphasize *either* control (thereby facilitating clarity of inference) *or* realism (i.e., correspondence with an ecologically representative state of affairs). Although I will question the either/or aspect of this dichotomy later, let us reexamine what it is we do when we opt for *control.*

Assuming our goal is truth, or theories about truth, we seek to make each laboratory experiment as tight as possible in order to ensure that each truth statement that results is as resistant as possible to the winds of criticism to which it inevitably will be exposed. The papers prepared for the symposium that gave rise to this volume express consensus that internal validity (i.e., clarity of inference) is the *sine qua non* of laboratory experimental research. We achieve this state by *control*. More specifically, the laboratory model urges us to seek *manipulative control*, since *creating* experimental variables leaves us in a better inferential position due to the selection biases and other contaminants that typically accompany "attribute" or "nonexperimental" variables (e.g., see Boring, 1969; Kerlinger, 1973). Certainly the laboratory experiment embodies an underlying logic that is elegant in its simplicity in addressing the issue of inference. Consistent with this view is the idea that laboratory experimentation is the ultimate arbiter of truth statements in science, since it is in that context that manipulative control and the precision it allows is maximized.

But consider the evaluation of a Decision Support System (DSS), Group DSS (GDSS), or other IS intervention in decision-making. In the real world, a DSS (or whatever) is implemented with considerable hoopla, and one or more reputations will probably be on the line for having recommended its adoption. In the lab, we will be concerned about placebo effects, and hence will do our best to nullify or neutralize such expectations. In the real world, individuals are employed for their particular expertise and their involvement with the implemented system reflects their power, position, and social relationships within the organization. In the lab, our *ceteris paribus* requirements will lead us to employ strangers wherever possible, and to randomly assign these participants to groups so as to equalize the distribution of competencies across conditions. In the real world, a system's performance has real consequences for the organization, and personal consequences for the careers of those involved with it. In the lab, there are token consequences at best, and the career of a research subject is a short one.

Now the laboratory is obviously a place apart from everyday life, but there is no compelling reason to believe that *every* aspect of life can or need be duplicated in order for research conducted in the laboratory to have some inferential utility. In the grander scheme of things, it probably does not matter much that the table in our lab is pine, while those in boardrooms are mahogany or oak. Any piece of research is a "simulation" insofar as it seems more relevant to express some elements than others (e.g., see Palys, 1978). But the factors cited in the previous paragraph would seem, intuitively at least, to be of a different order of magnitude. Nonetheless, the question is not whether there are differences between the laboratory and the world--of course there are--but rather whether these differences *make* a difference.

The laboratory has traditionally been seen as the optimal site in which to exercise the rigors of experimental control. The investigator can pluck phenomena from their earthly roots, scrutinize them in detail, and then replace them without damage. The assumption is that by decontextualizing the phenomenon one is engaging in an act of convenience, not destruction. If the laboratory is somehow a different place, its major distinguishing characteristic is that it provides a neutral and undistracting sanctuary for our investigations. Consistent with this view is the idea that our *ceteris paribus* dicta merely make for a more conservative test; if an intervention is shown to be efficacious in the absence of expectations and relationships and history of use, surely it will be so in spades when it arrives with bells and whistles in the organization. If there *are* some problems in extrapolating, we need only do some additional work to determine the interactional factors we should consider. This assumes that any questions about the relationship between the lab and the real world are basically questions of external validity; across what persons, settings, and times can the results be generalized?

Others would argue that the issue is not so much one of *external* validity as of *construct* validity. I sense that most of the authors of the other papers in this volume are sensitive to the idea that we do not so much *de*contextualize phenomena of interest when we pull them in to the lab as *re*contextualize them.[1] But in doing so we may create an entirely different phenomenon; in exerting control over a phenomenon, we may alter not only its location, but also its *meaning*, and hence activate different processes in response to it.

Is the act of "making a decision" essentially theoretically similar or qualitatively different in lab and organization? An extensive literature suggests that it might be quite different; indeed, Janis and Mann (1977) argue, in several hundred pages and scores of experiments, that it is critical to distinguish between "hot" (i.e., involving, consequential) and "cold" (i.e., uninvolving, inconsequential) decision-making processes because the dynamics of the two are qualitatively different. If so, one can question whether the jump back to reality is plausible or not. When no less a personage than Dickson (1988) has difficulty nominating "any empirical MIS research that has had any influence whatsoever on practice in the field," there is cause for concern. Of course, impact itself is not an adequate epistemological criterion. Perhaps persons in organizations simply do not understand the logic of our inquiry, and hence are not as compelled as we are by the results. Then, again, perhaps we don't understand well enough the relationship between our context and theirs. Those who practice laboratory experimentation are obliged to articulate the alchemic formulae that convert their observations into real-world gold and, ultimately, to answer skepticism of their research programs with the appropriate empirical demonstrations.

Constructing Explanations

Various authors in organization (e.g., Morgan, 1986) and other fields (e.g., see Manicas and Secord, 1983) have discussed empirical models in terms of how each represents a construction of reality rather than a literal rendition of it; reality reveals itself, but only through the means by which we allow it to do so. Any given method or methodology will reveal only a particular slice of truth, rather than *the* truth, which suggests that we might question what types of truths a method like laboratory experimentation can reveal. Stated a different way, we can ask whether and what systematic biases are imposed by the way we recontextualize phenomena in the lab. Pushed a step further, we can ask what the implications of these biases will be if the results of the laboratory research are used to make decisions in the real world.

Consider again the example in which our experimental objective is to evaluate the impact of an IS intervention such as a DSS or GDSS on decision-making. On what bases can their impact be evaluated? Can any *inherent* implications be identified? Bronfennbrenner (1977) argued some years ago, in another field, that we are out of luck if we believe that we can unearth *the* effects of social phenomena. His particular interest in that article was on the potential impact of divorce and daycare on child development. Bronfennbrenner was arguing not that divorce or daycare *have no* effects, but rather than any assessment we might make will reflect not only something of the shift in familial dynamics that occurs when family disruption intervenes or when children younger than five have considerable social/educational experience beyond the nuclear nest, but also the social context in which these social phenomena occur (e.g., societal views and stigma concerning divorce; availability and forms of daycare). It is the social context that gives the phenomena their meaning, and these meanings will mediate their effects. This in itself should give us pause when considering the costs of recontextualizing phenomena in the laboratory for closer scrutiny. Nonetheless, it implies little more than what I have already stated: phenomena in the laboratory and the dependent variables they influence may take on different meaning in the laboratory.

To take Bronfennbrenner (1977) seriously is to imply that contemporary assessment of any IS intervention implicitly reflects the world view of those in the organization with whom we identify. Theorists such as Foucault (e.g., 1970, 1972) argue that it is for precisely these reasons that our experiments contribute to a reaffirmation and empowering of *status quo* interests. For Foucault, science does not uncover *the* truth, for there is no *one* truth to be discovered. We may reveal *a* truth, but it will be only one of many that might be revealed. Research foci represent a choice among possible trajectories of investigation, and the trajectories chosen will reflect and reaffirm the interests of those who have the power and money to do the research. From this perspective, it is no accident that most IS research evaluates interventions such as DSS or GDSS in terms of their ability to generate "optimal" solutions, where optimal is defined in terms of the economic bottom line. We less frequently see "optimal" defined in terms of economic profit *and* ethical and social responsibility, for example. If the results of economic bottom line research results in the classification of interventions as "effective" or "ineffective" (i.e., "good" or "bad") on the basis of how well they serve strictly economic interests, it is a small step to see that those decision-*makers* who use the interventions best (i.e., the most "effective" decision-makers) will be implicitly (if not explicitly) defined as those who maximize such variables as market permeation and net profit.[2]

It is because of these concerns that the themes articulated in Benbasat (1988) and Rohrbaugh (1988), concerning the desirability of focusing on the *processes* of decision-making, and Rohrbaugh's (1988) additional attention to the values reflected in these processes, are particularly refreshing. Part of the effort in achieving precision and control in a laboratory experiment is related to establishing limits on and priorities regarding the tasks and measures that are developed. As Dickson (1988) affirms, conscientious development of such components represents a formidable task that requires persistence and devotion. One study and one researcher obviously cannot accomplish everything. At the same time, the effort required to impose "adequate" control should not, in the larger scheme of things, be an excuse for myopia.

Encouraging Control-Oriented Structures

Finally, concerns may be expressed regarding the extent to which manipulative control in the laboratory experiment may find a home in the real world. Though we expend much effort considering the ways attributes of reality may or may not influence the design and results of our laboratory experiments, few have commented on the ways the experiment may influence the reality we study. When we *do* demonstrate that an intervention in a laboratory experiment has an effect our audience considers desirable, we simultaneously offer an articulation (or operationalization) of the conditions under which those results were observed. To the extent that our results are produced in a setting that emphasizes manipulative control, passivity, and malleability among participants in accepting proffered definitions as given, and a centralizing authority who decides what is and isn't relevant, we may encourage the creation of similar conditions in the world (see also Argyris, 1975; Brandt, 1975). In this sense, our lack of influence in organizational settings (recall Dickson, 1988) is, perhaps, in fact merely a healthy sign that decision-making in the real world involves the consideration of more perspectives than we currently acknowledge in our quest for manipulative control.

A Moratorium on Laboratory Experimentation?

I've been harsh on the laboratory experiment thus far. The reader might take my concerns as a preamble to a call for a moratorium on laboratory experimentation while we debate its deleterious influence on humanity and mere guise as an avenue towards truth. But such are neither my beliefs nor my aspirations. I simply believe that we need to reemphasize some of the disjunctions, rather than concentrate on the continuities, between the laboratory and the world.

The bottom line for the laboratory experiment is its elegance and amenability as a venue for addressing *ceteris paribus* propositions. Within the context of the closed system that is the laboratory experiment, we *can* unearth particular effects and their simple or complex antecedents and, under those conditions, even show some potency in prediction. But, as Manicas and Secord (1983) argue, prediction in a statistical sense is not synonymous with understanding and, in any event, prediction in the open system that is the world is a whole new ball game from prediction in the closed, controlled situation of the lab.[3] Nonetheless, the authors grant that the laboratory has a significant role to play in delimiting what they call basic "structures" or "powers" possessed or influenced by phenomena of interest (see also Secord, 1986). Such knowledge can facilitate our ability to *explain* the impact of our interventions in the world (since we *can* assess the status of the independent variables of interest at the time we wish to understand), even though we may be unable, even in principle, to *predict* what particular impacts will occur.

My admonitions call for us to be not only more self-conscious and aware of the roles we play in constructing knowledge, but also more tolerant and encouraging of methodological heterogeneity in the IS field. Benbasat's (1988) call for research on process as well as outcomes, Rohrbaugh's (1988) considerations of value perspectives utilized in those processes, Zmud's and Hauser's (1988) call for field experimentation, DeSanctis's (1988) heterogeneous conceptualization of perspectives to be considered when undertaking laboratory research, Mason's (1988) call for triangulation and diversity, and Dickson's (1988) call for better laboratory research are all laudable. At the same time, there is expressed in all of these papers an apparent belief that "control" and "realism" are alternative ends of a continuum--that one must declare one's allegiance to one *or* the other in the context of any given study. This appears to be based on a perceived synonymy between "clarity of inference" and "manipulative control." I suggest that we distinguish instead between "manipulative control" and "analytical control," both of which offer the clarity of inference we seek within their respective limitations.

On Manipulative and Analytic Control

Donald T. Campbell has clearly been an influential figure in 20th century science. Our papers speak in a language he played a major role in developing, and most of us have referenced one or more of his classic texts (e.g., Campbell and Stanley, 1963; Cook and Campbell, 1979). Campbell's contributions lie largely in his explication and discussion of rules of inference, and in the vocabulary he provided regarding the logic that underlies experimentation, to which he drew our attention. Not knowing the content of his thoughts or the nature of his aspirations when he first told us about internal and external validity (et cetera) in the 1950s and 1960s, I can only admire the way his dimensionalizing of research paved the way for his subsequent discussions of quasi-experimentation in the 1960s and 1970s.[4]

Prior to Campbell, the laboratory was perceived as the only place in which one could do "real" science. Field research was nice and could yield fascinating and interesting findings, but the field itself was viewed as an inherently muddy place that yielded dirty data that were problematic from the perspective of inference. Campbell sought means by which these differences could be overcome to address problems of the day. Two of his papers, "Reforms as Experiments" (Campbell, 1969) and "Connecticut Speed Crackdown" (Campbell & Ross, 1968), were instrumental in pointing out what an important role social science could play in developing and evaluating social policy and interventions if we took our curious and experimenting attitude into the field. Equally important, he offered specific suggestions concerning how to go about it.

Overall, one of the most important things Campbell did was to admonish us to differentiate between the *trappings* of experimentation (e.g., clipboards, random assignment) and its underlying *objectives* (i.e., as a facilitator of inferences about cause, association, and influence). As one of Campbell's colleagues at Northwestern put it:

> The assumption of (quasi-experimental design) is that the experimental method has much broader applicability than its laboratory version suggests. . . . What is important is *not* (the) ability to manipulate and assign randomly, but the *ends* these procedures serve. . . . The problem then becomes one of providing the proper translation rules to get the social scientist out of the lab and into the "real world," while retaining some of the strong inference characteristics of the laboratory setting (Caporaso, 1973, pp. 6-7, emphasis and parenthetical phrases added).

Though the laboratory may be an appropriate site for some phenomena, and certain strategies might be useful in that setting, those that *do not* fit should neither be thrown on the bed of Procrustes nor ignored, but addressed in a different way as suggested by Campbell (1969).

> The advocated strategy in quasi-experimentation is not to throw up one's hands and refuse to use the evidence because of [a] lack of control, but rather to generate by informed criticism as many appropriate rival hypotheses as possible, and then to do the supplementary research . . . which would reflect on these rival hypotheses.

In sum, there are many ways to achieve control in order to facilitate causal inferences about social phenomena. The laboratory experiment is the embodiment of the idea that much can be accomplished when *manipulative* control is mobilized towards inferential ends, though the criticisms outlined earlier suggest limitations and potential liabilities in what can be accomplished. In contrast, the rationale that underlies Campbell's description of quasi-experimentation suggests that when manipulative control *per se* is impossible and/or undesirable, *analytic* control can offer an alternate route for achieving the same ends. Although Campbell's earlier articles betrayed the nomothetic bias that was prevalent in social science (i.e., that the ideal situation was one in which aggregated groups were compared in terms of central tendency), one might fruitfully extend his conceptualizations into the domains of case studies and everyday life. What is inference, after all, but an attempt to address Mill's three criteria: temporal precedence, association, and, most critically, *the elimination of rival plausible explanations*? In one way or another, one "makes a case," which involves showing why one's explanation of events is more compelling that anyone else's.

The Quasi-Experimentation of Everyday Life

Consideration of the "quasi-experimentation of everyday life" has utility for all prospective researchers, regardless of their methodological perspective. I offer the following, perhaps trite, example. Recently, after filling my car up with gas and entering the volume and cost of the purchase in my expense log, I noticed an appreciable change in expenses. I now filled up the tank every 12 to 14 days instead of once a week, and a tankful of gas cost about $18.00 instead of the $25.00 it had before. How splendid, I thought, that my new car was saving me even more than I anticipated after only a few months. This is clearly an implicit causal statement. I believed that buying a new car had caused my gas mileage to increase and expenses to decrease. Although I could have gone straight out and told the world about this profound conclusion, it was my obligation as a scientist to consider various rival explanations for this alleged link. Had my driving habits changed? Was I still driving the same distances? Had the cost of gas fluctuated? As it turned out, I had moved to a location that was closer to both the University and my son's daycare (my two most frequent destinations. What's more, I was also on sabbatical, spending more time at home finishing a book than going to the office. Although my new car *is* an improvement over the old one, the shift was not as extreme as my expense log alone suggested.

What is important here is not the example, which I have admitted is trivial,[5] but the logic that underlies it. I began by asserting a state of affairs regarding a matter of the real world that held some importance for me. It didn't matter that it was a conclusion that emerged (inductively) *after* driving my new car for awhile and observing the entries I had made in the log; I could just as easily have (deductively) hypothesized ahead of time that such an effect would result. In either case, I would be obliged to articulate my understanding of the relationships that exist among the various "relevant" factors and, indeed, to delineate why I found them to be relevant at all. The situation I recounted did not include representative samples of the population, experimental and control groups, or any attempt to randomly assign drivers to cars. I did not use a computer to analyze the data, and no test was performed to assess the statistical significance of the results.[6] Yet the logic underlying my example meets contemporary canons of science insofar as it attempts to determine whether the assertion I made could "account" for the data I suggested were relevant, and includes consideration of rival plausible explanations.

It is in this sense that I see Campbell's contributions going beyond the design concerns of laboratory experimenters or evaluation researchers. Such individuals will create, or adapt to, situations in which appropriate experimental and comparison groups will be created. But all they are doing is trying to account for data and assess the merits of competing explanations. These are the same objectives, it seems to me, that engage not only quantitative researchers who attempt to unearth mathematical models or perform laboratory experiments, but also qualitative researchers who eschew quantification and experimental design in order to describe the "culture" of a group via ethnography or the "meaning" of some social objects to some participants, or who wish to unearth the archaeology of knowledge à la Foucault.

Researchers raised with random assignment and MANOVA may feel unsettled by the uncertainty of applying experimentalist principles in the case study setting, in which qualitative researchers, ethnographers, ethnomethodologists, and the like have arrived at an occasionally worthy stereotype of inference by intuitive revelation. Campbell himself may have exacerbated the problem in his earlier writings, in which he explicitly labeled case studies "pre-scientific" and "of little inferential value." For example, in his influential monograph with Julian Stanley (Campbell and Stanley, 1963), Campbell caustically noted that

> as has been pointed out, such [case] studies have such a total absence of control as to be of almost no scientific value. . . . Such studies often involve tedious collection of specific detail, careful observation, testing, and the like, and in such instances involve the error of *misplaced precision*. . . . It seems well-nigh unethical at the present time to allow, as theses or dissertations . . . case studies of this nature. (pp. 6-7)

The main problem Campbell saw with such studies was that there were too many possible explanations and too few observations against which to assess the veracity of those explanations. I heard the classic example of this problem on a recent radio news program, which contained a story about a British gentleman who was celebrating his 111th birthday. As is usual in such interviews, the man was asked to what he attributed his longevity. Such a situation is problematic to the social scientist, in that it presents only one observation (i.e., that the man is 111) and a lifetime full of explanatory variables that could potentially account for that outcome. As Campbell later explained:

> The caricature of the single case study approach which I have had in mind consists of an observer who notes a single striking characteristic of a culture, and then has available all of the other differences on all other variables to search through in finding an explanation. . . . That he will find an "explanation" that seems to fit perfectly is

> inevitable, through his total lack of "degrees of freedom." (It is as though he were trying to fit two points of observation with a formula including a thousand adjustable terms, whereas in good science, we must have fewer terms in our formula than our data points). (Campbell, 1979b, p. 54)

Campbell later reconsidered both his earlier stereotype of qualitative, case study research and his evaluation of such studies. To wit:

> In past writings . . . I have spoken harshly of the single-occasion, single-setting (one shot) case study, not on the ground of its qualitative nature, but because it combined such a fewness of points of observation, and such a plethora of available causal concepts, that a spuriously perfect fit was almost certain. Recently, in a quixotic and ambivalent article (1975), I have recanted, reminding myself that such studies regularly contradict the prior expectations of the authors, and are convincing and informative to sceptics like me to a degree which my simple-minded rejection [did] not allow for. (Campbell, 1978, p. 201)

The turning point for Campbell was this: if case studies are so easily supportive of whatever explanation the researcher brings to the situation, why do so many qualitative researchers report being surprised, changing their beliefs, and revising their theories (see, for example, Becker, 1970; Campbell, 1978; 1979b)? Part of the reason for Campbell's change of heart was that he reconstrued case studies. He *had* seen the case study as *one* (collective) observation, as one would from the perspective of aggregate statistics. Later, he acknowledged that *myriad* observations are possible within the context of a given case study.

> While it is probable that many case studies professing or implying interpretation or explanation, or relating the case to theory, are guilty of [the faults he had outlined], it now seems to me clear that not all are, or need be, and that I have overlooked a major source of discipline. . . . In a case study done by an alert social scientist who has thorough location acquaintance, the theory he uses to explain the focal difference also generates predictions or expectations on dozens of other aspects of the culture, and he does not retain the theory unless most of these are also confirmed. In some sense, he has tested the theory with degrees of freedom coming from the multiple implications of any one theory. The process is a kind of pattern matching . . . in which there are many aspects of the pattern demanded by theory that are available for matching with his observations on the local setting. (Campbell, 1978, p. 57)

In sum, just about *any* theory can account for a *single* observation. The trick is to evaluate and develop the theory by looking at the *multiple* observations it implies and by constantly considering rival, plausible explanations that might account equally well for some or all of these observations. Once again, Campbell declares his allegiance to the *logic* of empirical inquiry.

The generic research process Campbell envisions is one in which a rigorous and self-critical scholar can use *any* source of information as a vehicle for generating or evaluating the multiple implications of theory.[7] To the extent that the scientist is attuned to such multiple implications, and systematic and forthright in evaluating the consistency of any given theory, "science is much better than ignorance and, on many topics, better than traditional wisdom. Our problem as methodologists is to define our course between the extremes of inert skepticism and naive credulity. *When a scientist argues that a given body of data corroborate a theory, invalidation of that claim comes in fact only from equally plausible or better explanations of those data*" (Campbell, 1978, p. 185; emphasis in original). We are led to the picture of science as a community of disputatious, questioning truth-seekers, whose role is to marry a critical approach with an anticipation of rival plausible explanations.

As I expressed it earlier in this paper, Campbell is arguing for "the quasi-experimentation of everyday life," and suggesting that "good" case studies are those that expend the effort to achieve thorough local knowledge, draw inferences through the process of offering explanations to account for observations, and eliminate rival, plausible explanations. This process has been articulated and developed most fully thus far by two former students of Campbell's--Louise Kidder (e.g., see 1981) and Paul Rosenblatt (e.g., see 1981)--who studied with both Campbell (initially a very quantitatively-oriented psychologist) and Howard Becker (initially a very qualitatively-oriented sociologist) at Northwestern University. Kidder (1981), for example, offers an integration of the quasi-experimentational framework with three different qualitative case studies, including Becker's (1963) classic study on "becoming a marihuana user." Becker (1979) offers a similarly qualitative account that is clearly framed within a "rival plausible explanations" perspective, which Rosenblatt (1981) argues has utility in ethnographic research. In terms of more elaborate treatments, Palys (forthcoming) offers an integration of qualitative and quantitative perspectives from within an overarching "rival plausible explanations" framework.

A Place for Qualitative Knowing

One of the incidentals of laboratory experimentation, emerging as it has from the positivist tradition, is its typically deductive approach. The theorist, whether on the basis of the extant literature, preliminary observation, or some form of intuitive revelation, puts together a preliminary theory about the prospective dynamics of a phenomenon of interest and proceeds to test it. This perspective places little emphasis on exploratory research, but may expend great effort in pilot studies or pre-tests devoted to developing or fine-tuning research instruments in accordance with a theoretically determined strategy. In contrast, inductivists have typically revelled in the exploratory phase, often appearing never to move beyond it. Their strength is that their accounts frequently reveal a phenomenological understanding from the perspective of participants that is often missing in deductivist portrayals.

By emphasizing a potential reconciliation of case study methods with experimentalist principles, a unification might be forged between the logical/rational discipline of the best of social science analysis and the phenomenological integrity of what have been considered more "qualitative" modes of research. We have distinguished for too long between the "subjective" understanding of participants and the "objective" aspirations of social scientists. Those who cling to this dichotomy are urged to reconsider, as Campbell (1979) did.

> Too often quantitative social scientists, under the influence of missionaries from logical positivism, presume that in true science, quantitative knowing replaces qualitative, common-sense knowing. The situation is in fact quite different. Rather, science depends upon qualitative common sense knowing even though at its best it goes beyond it. Science in the end contradicts some items of common sense, but it only does so by trusting the great bulk of the rest of common-sense knowledge. Such revision of common sense by science is akin to the revision of common sense by common sense which, paradoxically, can only be done by trusting more common-sense. (pp. 50-51)

This is *not* an invitation to indulge in naive credulity or gullibility. One can see the knowledge of science as "special" if for no other reason than that it emerges from a community of disputatious, quarrelsome truth-seekers who make the acquisition of such knowledge their business, and who believe in a constant interplay between theory and data. At the same time, our inquiry (in both laboratory and field) would be better informed by more extensive interaction with our milieu of interest, and by an appreciation of the phenomenology of those who lie within.

> When we get down to our own practical work, a plausible-rival-hypothesis approach is absolutely essential, and must for the most part be implemented by common-sense, humanistic, qualitative approaches. In programme evaluation, the details of programme implementation history, the site-specific wisdom, and the gossip about where the bodies are buried are all essential to interpreting the *quantitative* data.
>
> Qualitative knowing is absolutely essential as a prerequisite foundation for quantification in science. Without competence at the qualitative level, one's computer printout is misleading or meaningless.
>
> To rule out plausible rival hypotheses we need situation-specific wisdom. The lack of this knowledge (whether it be called ethnography, or programme history, or gossip) makes us incompetent estimators of programme impacts, turning out conclusions that are not only wrong, but are often wrong in socially destructive ways. (Campbell, 1984, pp. 30-34)

Emphasizing Inference Rather than Control

Contemporary lore has it that dependable knowledge is best generated when manipulative control is employed in the service of laboratory experimentation. The inferential capability that exists within the confines of a site is indeed considerable; its particular strengths lie in the analysis of "powers" or "structures" that underlie human behavior, its utility in establishing support for *ceteris paribus* principles that serve theoretical goals, its role as referee in establishing the plausibility of particular rival explanations in case study analysis, and its ability to stimulate possible realities that do not yet exist in the real world. But the laboratory has limitations, not the least of which is that science aspires to more than an understanding of laboratory behavior *per se*. Also of concern are the often ambiguous connections that exist between the laboratory and field contexts (e.g., see Tajfel, 1972), and the potentially negative implications of an exclusive reliance on empirical strategies that are manipulative and coercive by design (e.g., see Argyris, 1975; Brandt, 1975). Further limitations are discussed in other papers in this volume and elsewhere (e.g., Palys, 1978).

Fortunately, clarity of inference can also be achieved through nonmanipulative modes of control. Techniques emphasizing *analytic* control were first espoused by Campbell in the context of quasi-experimentation (e.g., see Cook and Campbell, 1979), though these were limited in that their execution presupposed an interest in and ability to deal with aggregate groups when making comparisons of strategic interest. I contend that the lessons extolled by Campbell in espousing quasi-experimentation can be applied fruitfully to case study designs. Acknowledging that case studies typically involve myriad observations changes the investigator's task to one of offering accounts that most adequately deal with the rival plausible explanations that are generated. Examples of this approach are most eloquently described in Kidder (1981) and Becker (1979), and further considered in Palys (forthcoming).

Instead of the traditional hierarchy of empirical "goodness" that tautologically reaffirms the desirability of manipulative control, the argument here puts the continuum on its side, and acknowledges that each of the alternatives listed employs some mixture of manipulative and analytic control in seeking possibilities for inference. It suggests that case studies that violate all the desiderata of experimental practice may nonetheless afford considerable inference capability *if* sufficient care has been taken to gain local knowledge and gather data that anticipate rival plausible explanations and conclusions.

Palys, Boyanowsky and Dutton (1984), for example, wished to evaluate, for police decision making and practice, implications associated with the implementation of a mobile computer system that allowed access to police-relevant databases. From a classic experimentalist perspective, it was the worst of all worlds--the opportunity to do the study came two years after the system was operational and no comparisons with other police departments were possible. The only alternative was a post-test only case-study design, which Campbell and Stanley (1963) described as involving "such a total absence of control as to be of almost no scientific value" (p. 6). Nonetheless, an extensive period of exploratory research, coupled with a multi-method assault involving overlapping and complementary sets of questionnaires, interviews, and observational and archival data, left the investigators feeling confident that they had achieved a rich understanding. Positive reactions from the police, the funding agency (a federal government department), and an audience of information systems analysts, and subsequent publication in *The Journal of Social Issues* suggest that the study had some scientific and applied merit.

Final Comments

This paper affirms that continuing allegiance to the *logic* underlying empirical inquiry, coupled with the phenomenological openness characteristic of case study approaches, holds significant potential for helping us to realize knowledge and theories that are not only accurate--in the sense of being veridical with respect to the reality they purport to describe--but also practical in their application. If this knowledge is generated as part of a program of inquiry that places at least as much emphasis on achieving inferential control through *analytic* as through *manipulative* means, then perhaps the implications of our research will enjoy the virtue of being more liberating than oppressive.

Endnotes

1. If not, Taijfel (1972) is quite forceful on this point.

2. Of course, these utilitarian shortcuts are not unique to business. In criminology, for example, the "bottom line" for correctional interventions is "recidivism." In one study that evaluated the impact of the introduction of a mobile radio data system on police decision-making, police expressed considerable

preference for equating "success" with "arrests," and less interest in considering social impacts, such as invasion of privacy issues and consequences for police-community interaction (see Palys, Boyanowsky, and Dutton, 1984). It is up to the investigator to be open to both the array of prospective impacts that might be considered and the various constituencies whose interests should be represented (e.g., see House, 1976).

3. This is because ceteris is never paribus in the world, and the complexities of everyday life leave us with an inability to predict the values that interactive factors will take on at a given moment. Because we cannot predict the context in which an event will occur, it is impossible for us to predict the status of the event itself at that time. The alternative for achieving predictability is to create similar control in the world, an alternative whose negative implications I addressed earlier.

4. If I have a complaint about Campbell's efforts at the time, it was that he chose to call this new domain *quasi*-experimentation, as in "almost"-experimentation, or "not quite as good as"-experimentation. My efforts here are largely an attempt to place the two on a more equal footing by showing their common inferential objectives and their respective advantages and limitations.

5. As an aside, it is interesting to consider why my example is best described as "trivial." Most directly it is due to the lack of a theory that contextualizes and provides meaning to my behaviour. You could probably care less that I bought a new car, or where I drive when I leave home, but it *would* be of interest to you if you were studying the impact of the automobile on 20th century society, or how single parents integrate child-rearing priorities with their other life roles.

6. Note, however, that part of my understanding of the phenomenon may well have been generated by the results of research that utilized such samples and procedures.

7. I hasten to add that Campbell is not suggesting that "anything goes," nor would he refrain from arguing that some situations or sources of information are "better" than others in terms of the confidence or inference they allow. But I believe he would also assert that no particular situation or type of situation is more inherently connected to "Truth," and that we should not avoid information purely because it is not packaged in the manner we might like. Whatever the situation, he would admonish us to self-consciously consider the inferences we wish to make in terms of all the rival, plausible explanations that are available or that we can generate. As an aside, he would also encourage each new generation of researchers to realize its obligation to generate a whole new set of rival plausible explanations, and to argue vehemently about "Truth."

References

Argyris, C. "Dangers in Applying Results from Experimental Social Psychology." *American Psychologist* 30 (1975): 469-85.

Becker, H. S. "Do photographs tell the truth?" In T. D. Cook and C. S. Reichardt (eds.). *Qualitative and Quantitative Methods in Evaluation Research*, 99-117. Beverly Hills, Calif.: Sage, 1979.

__________. *Sociological Work*. Chicago: Aldine, 1970.

__________. *Outsiders: Studies in the Sociology of Deviance*. New York: Free Press, 1963.

Benbasat, I. "Laboratory Experiments in Information System Studies With a Focus on Individuals: A Critical Appraisal." In I. Benbasat (ed.), *The Information Systems Research Challenge: Experimental Research Methods*. Boston: Harvard Business School, 1990.

Bhaskar, R. *Scientific Realism and Human Emancipation*. Bristol, England: Verso (New Left Books), 1986.

Boring, E. G. "Perspective: Artifact and Control." In R. Rosenthal and R. L. Rosnow (eds.), *Artifact in Behavioural Research*, 1-11. New York: Academic Press, 1969.

Brandt, L. L. "Scientific Psychology: What For?" *Canadian Psychological Review* 16 (1975): 23-34.

Bronfennbrenner, U. "Toward an experimental ecology of human development." *American Psychologist* 32 (1977): 513-31.

Campbell, D. T. "Can We Be Scientific in Applied Social Science?" In R. F. Conner, D. G. Altman, and C. Jackson (eds.), *Evaluation Studies Review Annual*, vol. 9, 26-48. Beverly Hills, Calif.: Sage, 1984.

_________. "A Tribal Model of the Social System Vehicle Carrying Scientific Knowledge." *Knowledge: Creation, Diffusion, Utilization* 1 (1979): 181-201.

_________. "'Degrees of Freedom' and the Case Study." In T. D. Cook and C. S. Reichardt (eds.), *Qualitative and Quantitative Methods in Evaluation Research*, 49-68. Beverly Hills, Calif.: Sage, 1979b.

_________. "Qualitative Knowing in Action Research." In M. Brenner, P. Marsh and M. Brenner (eds.), *The Social Contexts of Methods*, 184-209. London: Breem Helm, 1978.

_________. "'Degrees of Freedom' and the Case Study." *Comparative Political Studies* 8 (1975): 178-93.

_________. "Reforms as Experiments." *American Psychologist* 24 (1969): 409-29.

Campbell, D. T., and L. H. Ross. "The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis." *Law and Society* 3 (1968): 33-53.

Campbell, D. T., and J. C. Stanley. *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally, 1963.

Caporaso, J. A. "Quasi-Experimental Approaches to Social Science: Perspectives and Problems." In J. A. Caporaso and L. L. Roos, Jr. (eds.), *Quasi-Experimental Approaches: Testing Theory and Evaluating Policy.* Chicago: Northwestern University Press, 1973.

Cook, T. D., and D. T. Campbell. *Quasi-Experimentation.* Boston: Houghton Mifflin, 1979.

DeSanctis, G. "Small Group Research in Information Systems: Theory and Method." In I. Benbasat (ed.), *The Information Systems Research Challenge: Experimental Research Methods.* Boston: Harvard Business School, 1990.

Dickson, G. W. "A Programmatic Approach to Information Systems Research: An Experimentalist's View." In I. Benbasat, *The Information Systems Research Challenge: Experimental Research Methods.* Boston: Harvard Business School, 1990.

Foucault, M. *The Archaeology of Knowledge.* New York: Pantheon Books, 1972.

_________. *The Order of Things.* London: Tavistock, 1970.

House, E. R. "Justice in Evaluation." In G. V. Glass (ed.), *Evaluation Studies Review Annual*, vol. 1, 75-100. Beverly Hills, Calif.: Sage, 1976.

Janis, I. L., and L. Mann. *Decision-Making: A Psychological Analysis of Conflict, Choice, and Commitment.* New York: Free Press, 1977.

Kerlinger, F. N. *Foundations of Behavioral Research* (2nd ed.). New York: Holt, Rinehart and Winston, 1973.

Kidder, L. H. "Qualitative Research and Quasi-Experimental Frameworks." In M. B. Brewer and B. E. Collins (eds.), *Scientific Inquiry and the Social Sciences: A Volume in Honour of Donald T. Campbell*, 226-56. San Francisco: Jossey-Bass, 1981.

Manicas, P. T., and P. F. Secord. "Implications for Psychology of the New Philosophy of Science." *American Psychologist* 38 (1983): 399-413.

Mason, R. O. "MIS Experiments: A Pragmatic Perspective." In I. Benbasat (ed.), *The Information Systems Research Challenge: Experimental Research Methods.* Boston: Harvard Business School, 1990.

Morgan, G. *Images of Organization.* Beverly Hills, Calif.: Sage, 1986.

Palys, T. S. *Research Decisions: Quantitative and Qualitative Perspectives.* Toronto: Holt, Rinehart & Winston, (forthcoming).

________. "Simulation Methods and Social Psychology." *Journal for the Theory of Social Behavior* 8 (1978): 343-68.

Palys, T. S., E. O. Boyanowsky, and D. G. Dutton. "Mobile Data Access Terminals and Their Implications for Policing." *Journal of Social Issues* 40, no. 3 (1984): 113-27.

Rohrbaugh, J. Demonstration Experiments in Field Settings: Assessing the Process, Not the Outcome, of Group Decision Support." In I. Benbasat (ed.), *The Information Systems Research Challenge: Experimental Research Methods*. Boston: Harvard Business School, 1990.

Rosenblatt, P. C. "Ethnographic case studies." In M. B. Brewer and B. E. Collins (eds.), *Scientific Inquiry and Social Sciences: A Volume in Honour of Donald T. Campbell*, 194-225. San Francisco: Jossey-Bass, 1981.

Secord, P. F. "Explanation in the Social Sciences and in Life Situations." In D. W. Fiske and R. A. Shweder (eds.) *Metatheory in Social Science: Pluralisms and Subjectivities*, 197-221. Chicago: University of Chicago Press, 1986.

Tajfel, H. "Experiments in a Vacuum." In J. Israel and H. Tajfel (eds.), *The Context of Social Psychology*, 69-119. London: Academic Press, 1972.

Zmud, R. W., Olson, M., and R. Hauser. "Field Experimentation in MIS Research." In I. Benbasat (ed.), *The Information Systems Research Challenge: Experimental Research Methods*. Boston: Harvard Business School, 1990.