METHODOLOGY MATTERS: DOING RESEARCH IN THE BEHAVIORAL and SOCIAL SCIENCES

JOSEPH E. MCGRATH

PSYCHOLOGY, UNIVERSITY OF ILLINOIS, URBANA

23 OCTOBER 1994

◆

"Doing research" simply means the systematic use of some set of theoretical and empirical tools to try to increase our understanding of some set of phenomena or events. In the social and behavioral sciences, the phenomena of interest involve states and actions of human systems — of individuals, groups, organizations, and larger social entities — and the by-products of those actions.

The meaning of research evidence, in any area of science, is inherently tied to the means or methods by which that evidence was obtained. Hence, to understand empirical evidence, its meaning, and its limitations, requires that you understand the concepts and techniques on which that evidence is based.

This chapter is about some of the tools with which researchers in the social and behavioral sciences go about "doing" research. It raises some issues about strategy, tactics and operations. Especially, it points out some of the inherent limits, as well as the potential strengths, of various features of the research process by which behavioral and social scientists do research.

SOME BASIC FEATURES OF THE RESEARCH PROCESS

Doing research, in the behavioral and social sciences, always involves bringing together three sets of things:

(a) some *content* that is of interest,
(b) some *ideas* that give meaning to that content, and
(c) some *techniques or procedures* by means of which those ideas and contents can be studied.

For example, the contents of a study might involve the behavior of a jury, conversations in a family about buying a new car, the voting behavior of members of a community, littering in a park, courtship patterns in a small town, and so forth. The ideas might include the concept of attitudes, the notion that education affects political preferences, the concept of conformity, the hypothesis that groups whose members like one another perform tasks better than groups whose members do not like each other, and so forth. The techniques might include a questionnaire to assess individual attitudes, toward a car or a candidate or group mates; a set of procedures for observing family discussions about cars and money; a means to gather election returns; a plan to evaluate the quality of group task products; and so forth.

I will refer to these three sets of things more formally, as three distinct, though interrelated, domains:

(a) The *Substantive* domain, from which we draw contents that seem worthy of our study and attention;
(b) The *Conceptual* domain, from which we draw ideas that seem likely to give meaning to our results; and
(c) The *Methodological* domain, from which we draw techniques that seem useful in conducting that research.

Material in this chapter draws on ideas from a number of my earlier publications, including: McGrath, J.E., 1984; McGrath and Brinberg, 1984; Brinberg and McGrath, 1985; Runkel and McGrath, 1972; McGrath, Martin, and Kulka, 1982.

Furthermore, research always deals with several levels of phenomena: With *relations* between units or *elements* within a context or *embedding system*. The elements, relations, and embedding systems have different forms in each of the three domains [See Figure 1].

SUBSTANTIVE DOMAIN

In the substantive domain, I will call the units or elements Phenomena, and the relations among them Patterns of phenomena. These Phenomena, and Patterns of them, are the object of our study. For the behavioral and social sciences, the phenomena of interest involve the states and actions of some human systems —individuals, groups, organizations, communities, and the like — and the conditions and processes that give rise to and follow from those states and actions.

Another way to say this is to say that the behavioral and social sciences study "actors behaving toward objects in context". An example would be "an individual casting a vote in a county election". Another example would be "the number of units produced in the week of April 12th by group 32 of the production division of the Danville plant".

It must be understood that both "actors" and "context" here refer to human systems at any of a number of system levels —individual, group, community, organization, and so on. Different behavioral and social sciences specialize in the study of different human systems —that is, in the study of phenomena and patterns at different levels and of different kinds. The rest of this book presents material that illustrates many of the substantive phenomena and patterns that have been studied within the field of human-computer interaction.

CONCEPTUAL DOMAIN

For the social and behavioral sciences, the elements of interest in the conceptual domain are properties of the states and actions of those human systems that are the focus of study — properties of "actors behaving toward objects in context". These might include such familiar ideas as "attitude," "cohesiveness," "power," "social pressure," "status," as well as many others that are used in social and behavioral science research. *Relations* in the conceptual domain refer to any of a variety of possible ways in which two or more elements can be connected. Some of those ways are viewed as "causal" connections. Some are logical relations. Some simply are chronological relations. For example, two elements can be equal or unequal, they can be related linearly or non-linearly, one can be a necessary or sufficient cause of the other, one can include the other, the relation between them can be one way or reciprocal, and many more. Materials from the conceptual domain —properties, and relations among those properties — are the "ideas" that can give meaning to the phenomena and patterns that we study in the substantive domain.

METHODOLOGICAL DOMAIN

In the methodological domain, elements are methods. I will call the methods *Modes of Treatment* (of properties of phenomena). Modes of Treatment are different ways by which a researcher can deal with a particular feature of the human systems that are to be studied.

One set of such Modes of Treatment include various *techniques for measuring* some feature (that is, for assessing the state or magnitude of some property of some actors-behaving- in-context), so that the researcher can determine what value or level that feature has for each "case" to be studied. Measurement methods include such things as: a questionnaire, a rating scale, a personality test, instruments for observing and recording communications, techniques for assessing the quality of some products resulting from individual or group task performance, and the like. (More is said about kinds of measures near the end of this chapter.)

Modes of Treatment also include various *techniques for manipulating* some feature of a research situation (that is, some property of an actor-behavior-context). To carry out an experimental manipulation of a feature of the situation (sometimes referred to as "manipulating a variable") means making that feature have one particular predetermined value or level for certain "cases" to be studied and another specific preordained value or level for certain other "cases," so that the effect of differences in that property can be assessed by comparing those two sets of "cases." For example: You might want to study the effectiveness of a particular human-computer system by studying two sets of work groups, one set of groups working with that computer system and the other set doing the same tasks "manually". Social psychologists have tried to manipulate features of the systems they study by a number of techniques, such as:

(a) **giving instruction** to participants (e.g., trying to motivate them to try hard by telling them that there will be a valuable prize for the best product);

(b) **imposing constraints** on features of the environment (e.g., providing some participants with a particular software program that may help task performance, and providing other participants with a different or no program to carry out that function);

(c) **selecting materials** for use (e.g., trying to produce differences in task difficulty by giving some participants very difficult word problems to complete, and giving other participants easier problems of the same type);

(d) **giving feedback** about prior performances (e.g., trying to induce feelings of success or failure by telling some participants they did well, and telling others they did poorly, on a previous task);

(e) **using experimental confederates** (e.g., trying to establish different degrees of liking for fellow group members by having an experimental assistant who is pretending to be a normal participant work very hard in some groups and act indifferent in others).

(More is said about techniques for manipulating variables near the end of this chapter).

Modes of Treatment of variables also include a set of *techniques for controlling the impact of* various "extraneous" features of the situation — features that are important but that you are not going to measure or manipulate in a particular study. These include: techniques for *experimental control*, by which you make certain features take the same predetermined value for all cases in the study (e.g., study only 6-year-olds to control on age); techniques for *statistical control* by which you try to nullify the effects of variations in a given property within a study by "removing" those variations by statistical means; and techniques for *distributing the impact* of a number of features of the system and its context—without directly manipulating or controlling any one of them— so that such impact can be taken into account in interpretation of results. The most prominent means for distributing impact of a number of features is called *randomization*, and refers to procedures for the allocation of "cases" among various conditions within the study. These Modes for dealing with various features of the human systems to be studied — *measuring, manipulating, controlling* and *distributing impact* — are the basic sets of elements or "tools" by which social and behavioral scientists systematically gather empirical information.

Relations in the methodological domain have to do with the application of various *Comparison Techniques*. These are methods or techniques by means of which the researcher can assess relations among the values of two or more features of the human system under study. Such comparisons involve three sets of features of the systems under study: (a) the features that have been measured, and that are regarded as measures of the phenomena of interest (these are sometimes called "dependent variables"); (b) the features that have been measured or manipulated, and that are regarded as potential covariates of, or antecedents to, the phenomena of interest (these are sometimes called "independent variables"); and (c) all of the other features of the system that are relevant to the relations of interest (between dependent and independent variables), and that you have (or have failed to) control, or whose impact you have (or have failed to) distribute or otherwise take into account. Comparisons assess the covariation or association between the values of the first two sets (the dependent and independent variables), against the backdrop of the third set (i.e., other relevant features that were not studied directly but that nevertheless are a part of the meaning of results).

Most of the rest of this chapter will deal with features of the research process that emphasize the methodological domain, without much systematic consideration of either conceptual or substantive matters. The reader should keep in mind, though, that the research process, like a three-legged stool, always depends on materials from *all three* domains — content, ideas, and techniques.

	DOMAINS		
LEVELS	SUBSTANTIVE	CONCEPTUAL	METHOD-OLOGICAL
ELEMENTS	Phenomena	Properties	Modes of Treatment
RELATIONS	Patterns	Relations	Comparison Techniques
EMBEDDING SYSTEMS	Ongoing systems [e.g. human-computer systems	Conceptual Systems (e.g., field theory)	Research Strategies (e.g. laboratory experiment)

Figure 1: Domains and levels of concepts in behavioral and social science research.

RESEARCH METHODS AS OPPORTUNITIES AND LIMITATIONS

Methods are the tools —the instruments, techniques and procedures — by which a science gathers and analyzes information. Like tools in other domains, different methods can do different things. Each method should be regarded as offering potential opportunities not available by other means, but also as having inherent limitations. You cannot pound a nail if you don't have a hammer (or some functional equivalent). But if you do have a hammer, that hammer will not help you much if you need to cut a board in half. For that you need a saw (or the functional equivalent). And, of course, the saw would not have helped to drive the nail. So it is with the tools or methods of the social and behavioral sciences.

All research methods should be regarded as *bounded opportunities* to gain knowledge about some set of phenomena, some substantive domain. Knowledge in science is based on use of some combination of substance, concepts and methods. The meaning of that knowledge, and the confidence we can have in it, both are contingent on the methods by which it was obtained. All methods used to gather and to analyze evidence offer both opportunities not available with other methods, and limitations inherent in the use of those particular methods.

One good example of this dual nature of methods —both opportunities for gaining knowledge and limitations to that knowledge — is the widespread use of questionnaires and other forms of self-report in many areas of the social and behavioral sciences. On the one hand, self-report measures (questionnaires, interviews, rating scales, and the like) are a direct way, and sometimes the only apparent way, to get evidence about certain kinds of variables that are worthy of study: attitudes, feelings, memories, perceptions, anticipations, goals, values, and the like. On the other hand, such self-report measures have some serious flaws. For example: Respondents may try to appear competent, to be consistent, to answer in socially desirable ways, to please (or frustrate) the researcher. Sometimes respondents are reactive on such self-report measures without even being aware of it. These flaws limit, and potentially distort, the information that can be gained from such self-report measures. Other approaches to data collection, such as observation of visible behavior, may be difficult or impossible to use when studying particular kinds of variables. For example: How do you go about observing anxiety, or sadness, or some other emotion? In any case, while such methods may avoid some of the particular weaknesses of self-reports, those methods will have other different weaknesses.

Such is the dilemma of empirical science: All methods have inherent flaws, though each has certain potential advantages. You cannot avoid these flaws; but you can bring more than one approach, more than one method, to bear on each aspect of a problem. If you only use one method, there is no way to separate out the part that is the "true" measure of the concept in question from the part that reflects mainly the method itself. If you use multiple methods, carefully picked to have different strengths and weaknesses, the methods can add strength to one another by offsetting each other's weaknesses. Furthermore, if the outcomes of use of different methods are consistent, this way of proceding can add credibility to the resulting evidence. If the outcomes differ across different methods, then you can avoid misinterpretation of the resulting evidence by properly qualifying your conclusions.

This same general problem (that methods are inherently flawed, though each is flawed differently), and this same general prescription for dealing with it (by use of multiple methods), hold, as well, for research strategies, for comparison techniques and for research designs, all of which will be discussed subsequently in this chapter. For example, the research strategy called the laboratory experiment has some important strengths. It can permit precise measurement of effects resulting from deliberate manipulation of presumed causes, and therefore the drawing of strong inferences about cause-effect relations. But laboratory experiments also have some serious flaws. Researchers using laboratory experiments often greatly narrow the scope of the problem; they study it in artificial settings; and they are likely to use procedures and measures that make the situation seem even more artificial to the participants.

Several strategies that are alternatives to laboratory experiments are discussed later in this chapter. They include: field studies, sample surveys, and several others. Each of these other strategies offers different strengths, some of them offsetting the weaknesses of the laboratory; but each also has different inherent weaknesses, some of these being the very strengths of the laboratory strategy. No one strategy, used alone, is very useful; each of them is far too flawed. But again, the researcher needs to take advantage of multiple approaches. Usually, this cannot be done within a single study — often, the researcher must use a single strategy as a practical matter. But multiple strategies can be used over several studies of the same problem. The approaches need to be chosen so that the weaknesses of each strategy can be offset by the strengths of another. If we obtain consistent outcomes across studies using different strategies, we can be more confidant that those outcomes have to do with the phenomena we are studying, and not just with our methods.

To summarize:

(a) Methods enable but also limit evidence.
(b) All methods are valuable, but all have weaknesses or limitations.
(c) You can offset the different weaknesses of various methods by using multiple methods.
(d) You can choose such multiple methods so that they have patterned diversity; that is, so that strengths of some methods offset weaknesses of others.

Given these principles, it should be why it is not appropriate to ask whether any given study is flawless, and therefore to be believed (as in the query, "But is that study valid?"). Rather, we should ask whether the evidence from any given study is consistent with other evidence on the same problem, done by the same or other researchers using other strategies and other methods. If two sets of evidence based on different methods are consistent, both of those sets of evidence gain in credibility. If they are not consistent, that inconsistency raises doubts about the credibility of both sets. How much doubt we may have about the two sets of evidence depends on what else is known about the problem and the methods from still other studies. On the other hand, if all of the studies of a given problem have been based on the same methods, then that body of information is very much contingent on, and limited by, the flaws of those methods. Such a body of information must be regarded with some skepticism until you know whether it holds for a broader array of methods.

It should be noted here, though, that no one investigator is apt to be trained in the use of all methods, nor to have access to the resources needed for all of them. For example, some researchers have access to use of extensive and well designed laboratory facilities and are well trained in those methods but do not have ready access to the resources needed for a full scale sample survey, or for an elaborate field study. Other researchers may be in the reverse situation, with poor or no laboratory facilities but with excellent survey facilities and field study opportunities. What is crucial is not that a given researcher be able to use all methods on his or her research problem, but rather that the field as a whole make such use of diverse methods on each of its key problem areas. The fundamental principle, in behavioral and social science is that *credible empirical knowledge requires consistency or convergence of evidence across studies based on different methods.* These issues and their implications for behavioral and social science are discussed further in the parts of this chapter to follow, along with more detailed descriptions of strategies, comparison techniques, designs and methods.

RESEARCH STRATEGIES: CHOOSING A SETTING FOR A STUDY

Research evidence, in the social and behavioral sciences, always involves *somebody doing something, in some situation.* We can always ask about three facets: Who [which actors], what [which behaviors] and when and where [which contexts]. [The terms "actor", "behavior" and "context" are used here as technical terms with meanings somewhat different from ordinary usage. Actor refers to those human systems, at whatever level of aggregation (e.g., individuals, groups, organizations, communities) whose behavior is to be studied. Behavior refers to all aspects of the states and actions of those human systems that might be of interest for such study. Context refers to all the relevant temporal, locational and situational features of the "surround" within which those human systems are embedded.]

When you gather a batch of research evidence, you are always trying to maximize three desireable features or criteria:

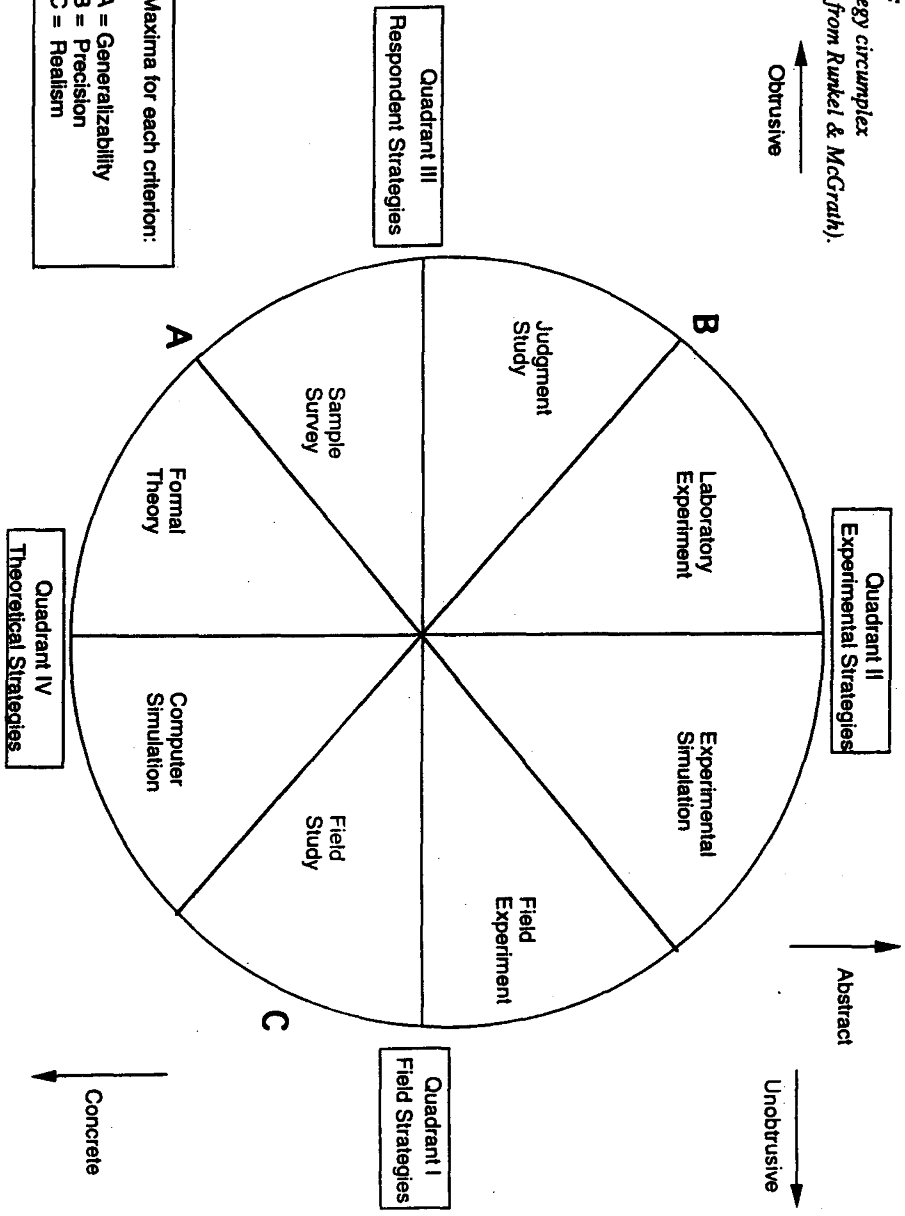
A. **Generalizability** of the evidence over the populations of Actors.
B. **Precision** of measurement of the behaviors that are being studied (and precision of control over extraneous factors that are not being studied).
C. **Realism** of the situation or Context within which the evidence is gathered, in relation to the contexts to which you want your evidence to apply.

Although you always want to maximize all three of these criteria, A, B and C simultaneously, you cannot do so. This is one fundamental dilemma of the research process. The very things you can do to increase one of these three features reduces one or both of the other two. For example: The things you can do to try to increase the precision with which you can measure behavior and control related variables (B) (for example, conducting a carefully controlled laboratory experiment) will intrude upon the situation and reduce its "naturalness" or realism (that is, reduce C), and will also reduce the range of actors (A) to whom the findings can be generalized. Conversely, the things you can do to try to keep high realism of context (C) (for example, conducting a field study in a natural situation) will reduce both the range of populations to which your results can be applied (A) and the precision of the information you generate (B). As a third example, the things you can do to try to establish a high degree of generalizability over actors (A) (for example, conducting a well-designed sample survey) will reduce realism (C) by obtaining the measures out of context, and will reduce precision (B) both by having measures of only a limited number of behaviors, and by failing to control or otherwise take into account extraneous factors that may affect results.

You can appreciate this dilemma better by examining some of the major research strategies used in the behavioral and social sciences. Figure 2 shows a set of eight alternative research strategies, or settings for gathering research information. In that figure, the eight strategies are shown as lying in a circular arrangement in relation to two underlying dimensions: the degree to which the setting used in the strategy is universal or abstract vs. particular or concrete; and the degree to which the strategy involves procedures that are obtrusive, vs. procedures that are unobtrusive, with respect to the ongoing human systems (the actor-behavior-context units) that are to be the object of study. The four strategies on the right side of the circle involve fairly concrete or particularistic settings; the four on the left side use fairly universal or abstract settings. The procedures used in the four strategies in the lower half of the circle can be fairly unobtrusive. The four strategies in the top half of the circle necessarily use procedures that are fairly obtrusive, that is, they disturb the ongoing human systems (the actor-behavior-context units) that are being studied.

Figure 2 also shows where, among the strategies, each of the three desired features, or criteria is at its maximum. Criterion A, generalizability with respect to the population of Actors, is potentially maximized in the sample survey and in formal theory. Criterion B, precision with respect to measurement and control of behaviors, is potentially at its maximum in the laboratory experiment and in judgment studies. Criterion C, realism of context, is potentially at its maximum in the field study. The geometry of figure 2-1 emphasizes the dilemma just discussed, namely: strategies that maximize one of these are far from the maximum point for the other two. The very same changes in research procedures that would let you move toward the maximum of any one of these criteria —A, B, or C —at the same time would move you away from the maximum point of the other two. *It is not possible, in principle, to maximize all three criteria simultaneously.* Thus, any one research strategy is limited in what it can achieve. Research done by any single strategy is flawed, although the various strategies are flawed in different ways.

Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



The eight strategies listed in Figure 2 are shown as four pairs, each occupying one quadrant of the circle. Quadrant I contains research strategies that involve observation of ongoing behavior systems under conditions as natural as possible. Quadrant II contains research strategies that are carried out in settings concocted for the purpose of the research. Quadrant III contains research strategies that involve gathering responses of participants under conditions in which the setting is muted or made moot. Quadrant IV contains research strategies that are theoretical, rather than empirical, in character. The two strategies in each of these quadrants will be described and illustrated briefly in the following paragraphs.

QUADRANT I: THE FIELD STRATEGIES

The two research strategies in quadrant I are the Field Study and the Field Experiment. In a field study, the researcher sets out to make direct observations of "natural", ongoing systems, while intruding on and disturbing those systems as little as possible. Much of the ethnographic work in cultural anthropology would exemplify this strategy, as would many field studies in sociology and many "case studies" of organizations.

A field experiment is a compromise strategy in which the researcher gives up some of the unobtrusiveness of the plain field study, in the interest of gaining more precision in the information resulting from the study. Typically, a field experiment also works within an ongoing natural system as unobtrusively as possible, except for intruding on that system by manipulating one major feature of that system. Field experiments use a manipulation of one important feature of the system in order to be able to assess the causal effects of the difference in that manipulated feature on other behaviors of the system. A number of studies in work organizations, such as the famous Western Electric or Hawthorne studies (Roesthlinger & Dickson, 1939), would exemplify the field experiment. Such studies introduce a major change in one feature of the organization (for example, a change in the formal communication structure), and study the changes that occur elsewhere in the organization subsequently. Sometimes such research also studies an unchanged but otherwise comparable organization, as a basis for comparison.

The essence of both of the strategies in quadrant I, the field study and the field experiment, is that the behavior system under study is "natural", in the sense that it would occur whether or not the researcher were there and whether or not it were being observed as part of a study. The two strategies of quadrant I differ in that the field study remains as unobtrusive as it can be (although no study is ever completely unobtrusive), at a cost in ability to make strong interpretations of resulting evidence; whereas the field experiment attempts to gain the ability to make stronger interpretations of some of the results (for example, that a behavior difference associated with the experimental manipulation may have been caused by the variables involved in that manipulation), but does so at a cost in obtrusiveness, hence in the naturalness or realism of the context.

QUADRANT II: THE EXPERIMENTAL STRATEGIES

The best known of the two strategies in quadrant II is the laboratory experiment. In that strategy, the investigator deliberately concocts a situation or behavior setting or context, defines the rules for its operation, and then induces some individuals or groups to enter the concocted system and engage in the behaviors called for by its rules and circumstances. In this way, the researcher is able to study the behaviors of interest with considerable precision (e.g., the investigator can be better prepared to measure certain behaviors because he or she can be confident about where and when those behaviors will occur), and to do so under conditions where many extraneous factors (that might be important but that are beyond the scope of the researcher's present interest) have been eliminated or brought under experimental control. The potential gain in precision in the measurement and control of behavior, which is the lure of the laboratory experiment, is paid for by increased obtrusiveness, hence reduced realism of context, and by a narrowing of the range of potential generalizability of results.

The other strategy of Quadrant II is the experimental simulation. In this strategy, the researcher attempts to achieve much of the precision and control of the laboratory experiment but to gain some of the realism (or apparent realism) of field studies. This is done by concocting a situation or behavior setting or context, as in the laboratory experiment, but making it as much like some class of actual behavior setting as possible.

One example would be research using ground-based flight simulators such as those used by both the U. S. Air Force and commercial airlines to train pilots for instrument flying. Another would be research that uses auto driving simulators like those sometimes used to train neophyte drivers. Still another would be research using military training exercises, or involving intra-squad practice games by an athletic team. Still another could be a monopoly game, or a stratgy game, or other similar board game, if they were used for research purposes and with some degree of control over "extraneous variables". Here, the key idea is that the researcher wants to create a system under his or her control, but at the same time have that system operate in a manner that simulates the operation of some particular class of naturally occurring system —the flight of airplanes, the steering of autos, the flow of "battle" in various sorts of two-sided combat or contests, or the operation of a "market" involving both strategic choices and chance factors.

The experimental simulation is a compromise strategy that attempts to retain the precision of the laboratory but at the same time to not give up so much realism of context. It risks introducing so much realism that precision of measurement and control are weakened, on the one hand, or retaining so much control that it becomes as "artificial" as the laboratory experiment on the other hand. An example of the former would be use of a military training exercise, in which the opposing "armies" are allowed to carry out any missions, anywhere and in any order, and thus make it impossible to observe and record the action for research purposes. An example of the latter would be to make such a "combat exercise" so stylized, and simplified in its flow — in the interest of good measurement and control — that all of the "realism" is nullified —that is, the system actually operating in the study does not function like the systems supposedly being simulated (that is, actual combat).

The two strategies in Quadrant II, in contrast to those of Quadrant I, involve concocted rather than natural settings. That is, the laboratory experiment and the experimental simulation are strategies that involve "actor-behavior-context" systems that would not exist at all were it not for the researcher's interest in doing the study. The distinction here is not between "real" and "unreal." The context of the laboratory experiment and the experimental simulation are certainly "real" for the participants once they are in the lab or simulation chamber; and the behaviors performed by the participants are certainly "real". Participants' behaviors are undoubtedly influenced by features of the experimental setting, but that is also the case in any other setting, natural or concocted. In fact, the exploration of such situational influences is, in large part, the point of behavioral and social science.

The distinction here, between the field research of Quadrant I and the experimental research of Quadrant II, has to do with whether the situation exists prior to and independent of the investigator, versus having been concocted by the researcher; and therefore whether the participants are taking part in it as an ongoing part of their lives or as part of a research endeavor. The issue is not one of reality, although much discussion of research strategies in the social sciences mistakenly treats it as such. Rather, the issue is one of motivation: Who has what stake in the behavior system under study.

Note that the difference between adjacent strategies are matters of degree. You can find "experimental simulations" (for example, varieties of strategy games) for which the task is so abstract that it becomes very close to a laboratory experiment. It also should be pointed out that few studies are "pure" examples of one strategy. These types of strategies represent a set of possibilities for carrying out research, rather than a description of concrete studies.

QUADRANT III: THE RESPONDENT STRATEGIES

In a sample survey, the investigator tries to obtain evidence that will permit him or her to estimate the distribution of some variables, and/or some relationships among them, within a specified population. This is done, typically, by careful sampling of actors from that population (thus potentially gaining a lot of generalizability, criterion A), and by systematically eliciting responses from those selected actors about the matters of interest. The many public opinion surveys on voting intentions, political preferences, buying intentions,and the like exemplify this strategy. While there is much emphasis on selection of sample, there is little opportunity for manipulation and/or control of variables and often little opportunity for much precision of measurement. Hence, this strategy is low on criterion B. And since the responses are gathered under conditions that make the behavior setting irrelevant, the question of realism of context is made moot (hence, this strategy is low on criterion C).

In a judgment study, the researcher concentrates on obtaining information about the properties of a certain set of stimulus materials, usually arranged so that they systematically reflect the properties of some broad stimulus domain. At the same time, such studies are usually done using "actors of convenience," so to speak. The focus of study is the set of properties of the stimulus materials, rather than some attributes of the respondents. Thus, studies using this strategy are often high on precision of measurement and control of both the stimulus materials and the responses (hence, high on criterion B). At the same time, they are often quite low on generalizability over population (hence, low on criterion A). Such studies are also usually done in a "neutral" behavior setting and with procedures that attempt to reduce or eliminate any properties of the behavior setting that might affect the judgments. Hence, they are low on criterion C. A good example of this strategy are studies in the area of psychology called psychophysics. These study the systematic relations between properties of the physical stimulus world and the psychological perception of those stimuli (e.g., the mapping of visible light and visual experience, of sound waves and auditory experience). Many "scaling" studies designed to explore the pattern of dimensions in some set of stimuli would also be examples of this strategy.

The two strategies of Quadrant III concentrate on the systematic gathering of responses of the participants to questions or stimuli formulated by the experimenter, in contrast to the observation of behaviors of the participants within an ongoing behavior system. Whereas the strategies of quadrant I focus on observation of behaviors within naturally occurring behavior settings, disturbed as little as possible by the research process, and the strategies of quadrant II focus on observation of behaviors within experimentally concocted behavior settings, the strategies of Quadrant III focus on observing behavior under conditions where the behavior setting is made irrelevant to the response. The sample survey does this by asking for responses that transcend the particular setting within which the responses are made. Questions such as how you intend to vote, whether you expect to buy a car next year, or how many children are living in your residence, are not related to the behavior setting within which a survey interview takes place (a doorstep, a living room, a shopping mall, an office). The judgment study makes the behavior setting irrelevant by attempting to neutralize or nullify the context of the behavior. To do this, the researcher attempts to mute, by "experimental controls," any of the features of the behavior setting that might "interfere with" the judgments being made. For example, psychophysical studies are usually done under "neutral" conditions of room temperature, lighting, chair comfort, and the like. The intent is to nullify any effects of the behavior setting or context on the judgments that are the topic of study.

The relation between the judgment study and the sample survey lies in their both emphasizing the behavior of some respondents in reaction to some stimulus materials, and deemphasizing the context within which those responses occur. The distinction between the judgment study and the sample survey has to do with two of their features: (a) whether the context is nullified by experimental controls or transcended by the nature of the responses elicited; and (b) whether the response of an individual to a stimulus is regarded as information about the stimulus (hence, a judgment study) or information about that respondent (hence, a sample survey).

QUADRANT IV: THE THEORETICAL STRATEGIES

Formal theory is a strategy that does not involve the gathering of any empirical observations (although it may be accompanied or preceded by much study of past empirical evidence). Rather, the researcher focuses on formulating general relations among a number of variables of interest. Generally, these relations —propositions, or hypotheses, or postulates —are intended to hold over some relatively broad range of populations. Hence this strategy is relatively high on the generalizability criterion A. At the same time, the formulation of theory in and of itself does not involve the operation of any concrete system (hence, it is relatively low on criterion C), nor does it involve the observation of any ongoing behavior (hence it is very low on criterion B). This strategy would be exemplified by any of the various general theories in behavioral and social sciences. Such theories are based on earlier empirical evidence (it is to be hoped), and they often lead to subsequent empirical studies. But the representation of the theory itself is not empirical — that is, it does not involve any "actors behaving in context".

The other non-empirical strategy in Quadrant IV is called Computer Simulation. It is like the experimental simulation strategy of quadrant II in that it is an attempt to model some particular kind of real-world system — a battle, a market, an aircraft in flight. But it is quite different from that strategy too. The computer simulation is a complete and closed system that models the operation of the concrete system without any behavior by any system participants. It does this because the researcher has designed a model that is complete and logically closed. All the important components of the system are specified by the investigator, and so are all of the relations among those components. Then, when the researcher starts a "run" of the system, all that ensues is the relatively predictable resultant of features built into the system. Such models are based on behavior in the sense that they must have all behavior parameters specified in advance, and this is often done on the basis of evidence from prior empirical research (at least for the parts about which the investigator has such past research available). But *no new behavior transpires during the run of the simulation.* And the "behavioral outcomes" that the simulation "shows" have the logical status of *predictions* from the theory that the researcher built into the model, rather than the status of behaviors occurring in nature independent of the control of the investigator. So this strategy is very low on criterion B. At the same time, it is potentially high on criterion C, in the sense that it is an attempt to model some concrete class of real world system (such as the geophysical processes going on in connection with the eruption of Mount St. Helens, or the prediction of the outcome of next year's Superbowl). But a computer simulation is designed to model some particular class of system; so the model is likely to have little generality over populations of actors or situations —or, more accurately, the question of generality over populations is moot. So this strategy is low on criterion A.

The two strategies of Quadrant IV are different in kind from the other six, but in a sense the pairs in each quadrant also differ in kind from those of the other quadrants. The inclusion of the two non-empirical strategies in this context is valuable for at least two reasons. First, the two theoretical strategies are related to the empirical strategies in several ways indicated in the diagram and in the preceding discussion. Second, the inclusion of these two strategies reminds us of the importance of the theoretical side of the research process. One major limitation of social psychology during much of its history (about 100 years) has been a reluctance to give full emphasis to the theoretical basis that is a necessary underpinning for any science. Inclusion of these two strategies also gives us the opportunity to note that one of the more powerful general strategies for research, and one that involves the use of multiple strategies on the same problem, is the simultaneous use of one of the theoretical strategies (say, the formulation of a general theory) and one of the empirical strategies (for example, a laboratory experiment).

SOME STRATEGIC ISSUES

Within the other chapters of this book, you will find studies done by all or most of the strategies discussed here. You should view that substantive material with two strategic issues in mind:

First, each strategy has certain inherent weaknesses, although each also has certain potential strengths. These weaknesses and strengths become part of the meaning of any evidence gathered with those strategies. So, an adequate interpretation of the available evidence on any given topic or problem should take those methodological strengths and weaknesses into account. The first strategy issue you are encouraged to address, in relation to material presented in the rest of this book, therefore is: Does the material, as presented, properly reckon with the strengths and weaknesses of the research strategies it encompasses?

Second, since all strategies are flawed, but flawed in different ways, to gain knowledge with confidence requires that more than one strategy —carefully selected so as to complement each other in their strengths and weaknesses— be used in relation to any given problem. While all of the research strategies discussed here are quite frequently used, there seems to be a tendency to use certain strategies for research on some problems or topics and other strategies for work on other problems or topics, but not to use multiple strategies on the same problem. The second strategic issue you are encouraged to address, with regard to the substantive material presented in the rest of this book, therefore is: To what extent is the research evidence on each problem or topic based on use of only a single research strategy, and therefore limited by the weaknesses of that strategy; and to what extent is that body of evidence based on use of multiple, complementary strategies, with agreement or convergence among the findings attained via the different strategies? The answers to those two issues are important indicators of how much the study of human-computer interaction has become a viable science with a cumulative body of credibly interpretable evidence.

STUDY DESIGN, COMPARISON TECHNIQUES, AND VALIDITY

In every empirical study, observations must be gathered, those observations must be aggregated and partitioned, and some comparisons must be made within that set of data. The comparisons to be made are the heart of the research. They reflect the relations that are the central focus of study. What comparisons are to be made in a given case depends on: (a) what has been included in the study at the element level from all three domains (what phenomena, what properties, what modes for treatment of variables have been used); (b) what systems are being worked from in all three domains (what substantive system is being studied, what conceptual paradigm is being used, what research strategy is being drawn upon); (c) what conceptual relations have been posited for the patterns of phenomena of interest (e.g., that a certain pair of properties, X and Y, are causally