Process tracing methods for the study of cognition outside the experimental psychology laboratory

Chapter · January 1993

David D Woods
The Ohio State University

Chapter 13

Process-Tracing Methods for The Study of Cognition Outside of the Experimental Psychology Laboratory*

David D. Woods
Cognitive Systems Engineering Laboratory
The Ohio State University

> to be successful in unlocking the doors concealing nature's secrets, a person must have ingenuity. If he does not have the key for the lock, he must not hesitate to pick it, to climb in a window, or even kick in a panel. If he succeeds, it is more by ingenuity and determination than by method. (Hildebrand, 1957, p. 26)

STUDIES OF COMPLEX BEHAVIORAL SITUATIONS

I will approach the topic of research methods for the study of human decision making and problem solving outside the usual psychology laboratory, that is, human cognition as it occurs in its natural setting (or situated cognition), in terms of a dichotomy between studying human cognition in complex, rich, multifaceted settings versus simplified, spartan, single-factor settings. Each setting offers different potential for insight into human behavior, and each is subject to different kinds of biases or sources of uncertainty.[1]

* Research support was provided by the Aerospace Human Factors Research Division of the NASA Ames Research Center under Grant NCC2-592. Everett Palmer served as technical monitor. I would like to thank the many colleagues who provided many useful comments and critiques. There were too many to mention all by name, but I am extremely fortunate that so many would help so much to improve this work.

[1] This distinction, I hope, focuses attention on constructive debate about research methods rather than the destructive debates framed in terms of ecological versus laboratory, applied versus basic or other similar dichotomies (e.g., Banaji & Crowder, 1989). Similarly, I hope to avoid a mere sermon on virtues associated with more ecologically valid studies.

This distinction defines the challenge that I will examine—how does one achieve valid, generalizable results when examining complex behavioral situations.

Why Study Cognition Outside the Laboratory?

Studying human behavior in complex situations is extremely difficult. So a dominant tactic people use to manage complexity in this as is in other difficult tasks is to bound the situation under consideration by focusing on one isolated aspect, cut off from the simultaneous function of other aspects with which it is normally integrated (Bartlett, 1932). Thus, in this spartan or austere approach, one might address only a single time slice of a dynamic process, or only a subset of the interconnections between parts of a highly coupled world. The strategy is to understand one variable or one subprocess at a time and then to consider how to put the pieces back together again (e.g., tachistoscopic-based research in perception).

However, this spartan, elemental approach is just one research strategy, which has limits. It is not clear with the spartan strategy whether the relevant aspects of the whole target situation have been captured in a test situation that extirpates or paralyzes most aspects of the whole. The importance of various parts of the problem-solving process may be underestimated, for example, "predecisional processes" (e.g., Friedman, Howell, & Jensen, 1985). Some aspects of problem solving may emerge only when more complex situations are examined directly. For example, there has been a great deal of research on human fault diagnosis. However, this work has almost exclusively addressed static devices with single faults, which is an oversimplification of many diagnostic situations where human problem solvers must cope with the possibility of multiple failures, misleading signals, interacting disturbances (e.g., Woods, Roth, & Pople, 1987). The assumption of a static situation has resulted in the failure to see a critical part of dynamic problem solving—*the disturbance management* cognitive activity (Woods, 1988) where managing the process to cope with the consequences of faults, that is, disturbances, goes on in parallel with and interacts with fault diagnosis.

Results from previous research (much of it of the spartan variety, ironically) also point to the need to investigate more complex behavioral situations. If thinking is a skill, as the Bartlett/Craik tradition in information processing holds, then we can study it in those who possess the skill. If skilled thinking is grounded in particular contexts (fields of knowledge-in-use), then the phenomenon of interest exists to be studied in the exercise of the skill in these contexts. If, as a broad assessment of studies of judgment and decision making indicates (e.g., Hogarth, 1986), strategies for judgment and choice are task-contingent, then we need to understand the role of that context in information-processing strategies, rather than always eliminate it from the task as in the spartan strategy.

These examples show the danger of falling into the psychologist's fallacy described by William James (1890), where the psychologist's reality is confused with the psychological reality of the human practitioner in his or her problem-solving world. One cannot simply assume that the experimenter's representation of the task is the same as the participant's representation (Cole & Scribner, 1974; Hutchins, 1980; Lave, 1988). The burden is on the experimenter (of either the spartan or complex persuasion) to determine or measure the participant's representation of the task so as to be able to interpret and generalize the observed behavioral results.

In reducing the target behavioral situation to a tractable laboratory or desktop world in search of precise results, we run the risk of eliminating the critical features of the world that drive behavior. But there seems to be an implicit assumption that researchers must suffer this risk, because the alternative is to study complex situations directly, which means all experimental control or focusing is necessarily abandoned. The familiar story of the drunk searching for the lost keys where the streetlamp casts light, rather than where the keys were lost, points out that both sources of uncertainty need to be overcome for effective behavioral science research. This chapter explores some of the ways that methodological lamps can be directed at complex behavioral situations.

Representativeness and the Mapping between Test and Target Behavioral Situations

Instead of focusing on the elemental, spartan strategy of throwing away complexity to achieve tractability, the appropriate criterion for creating tractable study situations is establishing a *mapping* between the *test* behavioral situation (where one is observing and measuring behavior) and the *target* behavioral situation one wishes to understand or generalize to.

We are often stuck with the appellation *real world* to distinguish research directed at complex settings from more spartan laboratory work. However, this terminology obscures the more relevant concept of the *representativeness* of one's experimental situation, that is, the relationship between the specific situation that is under study with respect to the class of situations that is the target of the study (Brunswik, 1956; Hammond, 1986). It is this test behavioral situation/target behavioral situation relationship that is the critical prerequisite for building a generalizable behavioral science research base. This is true whether the generalization is from a spartan situation to a complex one or from one complex situation to another complex one.

This point of view has several implications for research methods. First, there is the need for a much better understanding of the kinds of problems to be solved. To achieve representativeness we need to be able to analyze the "formal" characteristics of the situations towards which we want to generalize and map those characteristics into specific features of the test situation where we will actually observe and measure behavior. "Without a theory of task systems, however, it is impossible to know how to apply or to generalize the results of any given study" (Hammond, 1988, p. 3; cf. also Hogarth, 1986). For example, research results on one-shot decisions (a test situation frequently used in the laboratory) are of limited relevance for dynamic situations where information comes in over time and the problem solver must also decide when to act (e..g, Kleinmuntz & Thomas, 1987; Woods, Roth, & Pople, 1987). However, the optional-stopping decision problem does capture some of the characteristics of dynamically evolving situations and therefore can provide results transportable to other situations that possess these characteristics (Schwartz & Howell, 1985).

When one studies complex behavioral situations, the multifaceted nature of the setting creates the methodological and theoretical problem of deciding what counts as effective stimuli out of the total array. Note how this is analogous to the development of the area of ecological perception, in contrast to the "minimalist" research strategy in perception, where the commitment to studying more complex perceptual situations led to the need for a better understanding of the stimulus world—ecological physics. The effective stimuli in a multifaceted situation can be characterized, and the means is a semantic and pragmatic analysis of environment–cognitive agent relationships with respect to the goals/resources of the agent and the demands/constraints in the environment. For example, Bartlett (1932, p. 4) comments:

> We may consider the old and familiar illustration of the landscape artist, the naturalist and the geologist who walk in the country together. The one is said to notice and recall beauty of scenery, the other details of flora and fauna, and the third the formations of soils and rocks. In this case, no doubt, the stimuli being selected in each instance from what is present, are different for each observer, and obviously the records made in recall are different also. Nevertheless, the different reactions have a

uniformity of determination, and in each case spring from established interests.

This type of model of the characteristics of the setting relative to a practitioner's interests is what the person–machine system community tends to call the cognitive task analysis or cognitive work analysis (Rasmussen, 1986; Woods, 1988; Mitchell & Miller, 1986), and it is critical for characterizing the demands of problem-solving worlds (e.g., Roth & Woods, 1989).

In addition, the concept of representativeness points to more perception-like research programs where discovery of phenomena and demonstration of control of phenomena is the primary goal, rather than hypothesis testing per se. This style of research is oriented to characterize phenomena, to explore the factors that produce and modify phenomena, and to develop models that may capture the underlying psychological mechanisms.

PROCESS-TRACING METHODOLOGIES

This part of the chapter addresses various techniques that have been used to study complex behavioral situations. This section necessarily covers broad categories of techniques because of the extreme diversity of methods used, and because virtually every new, major "naturalistic" study includes some methodological innovation. However, the main focus of the discussion here will be protocol analysis or process-tracing methods. Another class of techniques has been developed for measuring the organization of knowledge possessed by an individual, especially expert practitioners in some domain. These techniques use classification and scaling methods derived from theory and research on semantic memory and concept formation to assess the kinds of categories by which the practitioner parses the domain and the relationships between these categories (e.g., semantic networks). To begin to examine these techniques in more detail see Chi, Feltovich, and Glaser (1981) and Cooke and McDonald (1987), as well as the work based on the classic deGroot memory paradigm.

The term *protocol analysis* has been used in a variety of ways, creating some confusion. I prefer to use the label *process-tracing methodologies,* which is more descriptive of the basic character of a wide variety of techniques (e.g., Ford et al., 1989). The goal in these methods is to map out how the incident unfolded including available cues, those cues actually noted by participants, and participants' interpretation in both the immediate and in the larger institutional and professional contexts. This is called a process-tracing or protocol analysis method because it focuses on how a given outcome came about.[2]

The specific techniques within this family are all oriented towards *externalizing internal processes or producing external signs that support inferences about internal workings*. To this end there are innumerable techniques and variants that have been used and that will be invented for tomorrow's study (Kato, 1986). One common technique is to transform the target behavioral situation into one that requires cooperation between two people (for examples, see Miyake, 1986; Suchman, 1987). This manipulation generates protocols based on verbal behavior that occur as part of the natural task behavior, rather than having participants produce concurrent verbal reports as an additional task. Note how the choice of manipulation in the test situation—change the task to a cooperative one versus ask for concurrent verbal reports—represents a tradeoff in sources of uncertainty about how the externalized cues relate to internal processes.

Another technique has been called *withheld information* (cf., as examples, Duncan, 1981; Johnson, Payne & Bettman, 1988). This technique is designed to externalize data acquisition and monitoring behavior during a problem-solving episode. Rather than having the entire set of data or data channels available in parallel for the study participant to examine, the experimenter withholds state information until the problem solver explicitly requests a specific datum. This allows the experimenter to watch the evolving process of data search to characterize the state of the underlying process (variable x is not behaving as expected), what knowledge is activated based on each observation (e.g., generating an hypothesis which might account for observed anamolous behavior in a device), which in turn directs new explorations of the data field (Woods, Roth, & Pople, 1987). Thus, this technique is particularly suited for making portions of the perceptual cycle more observable. However, the manipulation which produces observable signs of internal cognitive processing also produces a mismatch between the test behavioral situation and the target situation, and a source of uncertainty in data interpretation. In this case, the withheld information technique is not capable of supporting insight into processes associated with data-driven monitoring for new events and changes in the state of the underlying device, and the role of physically parallel data representations in those processes.

Note how there is a tradeoff where techniques that help externalize

---

[2] The increasing interest in including process as well as outcome in studies of human decision making can be seen in the formation of the European Group for Process Tracing Studies of Decision Making; cf. Montgomery and Svenson (1989a).

and make observable internal processes also can introduce distortions that reduce the accuracy of the mapping between test and target behavioral situations. Basically, there are two challenges to validity that the investigator attempts to cope with in the design of a process-tracing study (Russo, Johnson, & Stephens, 1989): (a) does the assessment technique change the primary processes that are the target of study (and how does it change those), and (b) to what degree do the data accurately reflect the underlying cognitive activities minimizing omission of important aspects, intrusions of irrelevant features, or distortions of the actual processes (and what checks on these sources of invalidity are included in the design)?

Verbal Reports

Process-tracing techniques primarily use data from verbal reports or from records of problem-solver behavior to build protocols that describe the sequence of information flow and knowledge activation. In addition, process-tracing techniques can be used to address critical incidents that have already occurred in retrospective analyses.

One type of process tracing is based on data derived from verbal reports made by study participants *about their own* process of solving the problem posed. This is not to be confused with verbal behavior, that is, task-related behavior that happens to be verbal, such as verbal communication in multiperson situations.

The debate about the validity of verbal data or about the importance of unverbalizable components of performance is large and ongoing (cf., e.g., Nisbett & Wilson, 1977; Ericsson & Simon, 1980; Berry & Broadbent, 1984). Verbal reports are just another kind of data which can be interpreted to provide insight and are subject to a number of dangers (e.g., Bainbridge, 1979; Praetorius & Duncan, 1988; Russo et al., 1989). Overall, there is agreement that it is critical to avoid and guard against verbal reports as introspections where the study participants analyze their own processes or behavior.

Techniques for verbalization include:

- Thinking-aloud protocols, where participants are instructed to think aloud as they work on the problem, that is, externalize the contents of working memory in the Ericsson and Simon view (cf. Russo et al., 1989).
- Retrospective verbal reports where participants solve the problems posed and afterwards provide a commentary about what they were thinking about at various points, for example, debriefing sessions (cf. Fidler, 1983).

- Cued retrospective verbal reports where participants comment after the problem-solving session but where the verbal report is cued to a record of their behavior during the case, for example, videotape (cf. Leplat & Hoc, 1981; Hoc & Leplat, 1983).

Behavioral Protocols

This technique has been developed in the context of domains where there is some underlying engineered or physiological process (e.g., Hollnagel, Pederson, & Rasmussen, 1981; Woods, O'Brien, & Hanes, 1987; Johnson, Zualkernan, & Garber, 1987). The human role is to manage that process in the face of disturbances produced by faults (domains where this occurs include aircraft flightdecks, managing space missions, nuclear power plant control rooms, air traffic control, managing patient physiology during surgery).

Rather than focus exclusively on participant verbalizations, behavioral protocols are built from a variety of data sources about the behavior of the people in relation to changes in the underlying process over time. Data sources include (a) direct observation of participant behavior, (b) traces of data acquisition sequences, (c) traces of actions taken on the underlying process, (d) records of the dynamic behavior of critical process variables, (e) records of verbal communication among team members or via formal communication media, (f) verbal reports made following the performance, and (g) commentaries on their behavior made by other domain knowledgeable observers. Data from all of these sources are correlated and combined to produce a record of participant data acquisition, situation assessment, knowledge activation, expectations, intentions, and actions as the case unfolds over time (cf. Woods, O'Brien, & Hanes, 1987, for several examples from simulated and actual nuclear power incidents). Note how different types of verbal behavior and verbal reports may contribute to the available lines of evidence. In behavioral protocol analysis the experimenter actively cross references the different lines of evidence in order to establish a trace of participant behavior and cognitive activities. This cross-checking and integration can help support the validity of the data with respect to participant cognitive activities at some level of analysis.

Typically, a major activity in behavioral protocols (as in any protocol analysis) is using knowledge of the domain to fill in gaps between observables. The raw data records may establish what a person did and in what context (what actions and signals had preceeded it, what did the team say to each other before the action was taken, etc.); however, these observables do not directly establish the person's intentions or situation assessment. But in fact one can establish what these are likely to be in most cases for behavioral situations where the human role is supervisory control (Rubin, Jones, & Mitchell, 1988). This is because there is usually only one interpretation or very few possible alternatives, given domain knowledge and the assumption of limited rationality; that is, human behavior is assumed to be the result of limits on rationality—people behave reasonably given their knowledge, their objectives, their point of view, limited resources (e.g., time or workload), the demands of the situation (Reason & Mycielska, 1982; Rasmussen, 1986; Woods & Roth, 1988). The assumption of limited rationality is used to understand human behavior *from the point of view of the person in the situation* rather than from the point of view of an omniscient observer, in order to reduce difficulties caused by hindsight bias, including the psychologist's fallacy. This is a fundamental objective of a process-tracing analysis—to lay out the problem-solving episode from the point of view of the people in the problem. The methodological tactics are selected or created to understand and represent the point of view of practitioners in the problem (either the specific people in a specific incident or the general view of the practitioner population).

The basic target to be achieved in a behavioral protocol analysis is tracing/understanding the evolution of the state of the underlying process or device *in parallel with* the human agents' state of understanding (situation assessment), intentions, and activities in managing the process. This means understanding discourse and action, data gathering, and situation assessment in relation to an external device/process—the referent world—which is itself changing both as a function of new events (e.g., faults) and corrective actions. It also includes understanding how changes in the referent domain activate new knowledge and trigger/shift lines of reasoning in the agents managing the process. Of course, signals can be missed or misinterpreted, knowledge can be buggy, relevant knowledge may not be activated, all of which can lead to mismatches between the agents' perception of the situation and the actual state of affairs and to erroneous actions (Woods, Roth, & Pople, 1987). Identifying these mismatches, and the circumstances that led to them, is the goal of a successful behavioral protocol analysis.

One useful technique to support behavioral protocol analysis is to use domain experts other than the study participants to observe episodes or review data records to help establish the participant's intentions and interpretations. This domain knowledge functions as a background for interpreting the behavior of study participants and may need to be captured more completely and formally as a cognitive task and work analysis (this may be a prerequisite for being able to build behavioral protocols). Mitchell and her colleagues (Rubin et al., 1988) have taken the next step and used the results of a cognitive task analysis as the knowledge base for a cognitive simulation that can fill in or delimit the intentions that would account for observable practitioner behaviors as an incident unfolds.

Retrospective Analyses of Critical Incidents

Retrospective analyses refer to cases where the incident of interest has already occurred (i.e., the classic critical incident technique in human factors). However, some data are available about the incident itself—one can review flight recorder transcripts, interview the participants in the incident after the fact, explore the context prior to the incident. This type of study is particularly important in investigations of human error in rarely occurring but very high-consequence situations where it is difficult to create the situation of interest (cf. Pew, Miller, & Feehrer, 1981; Woods, O'Brien, & Hanes, 1987; Klein, 1989b, for examples of retrospective analyses of human problem solving). A broad assessment of the methodological status and challenges of retrospective analyses of decision making is needed, but I will be content to sketch out some of the important issues here.

The assumption of limited rationality is important in applying a process-tracing method to past incidents. The participant's reports and other data records specify a sequence of activities. Clearly definable events, such as specific observations and actions, are used as starting points. The investigators use the participant's reports and the knowledge of other domain experts to interpolate the kinds of knowledge activated and utilized that would make this sequence of cues and actions rational from the point of view of limited cognitive agents. In other words, one reconstructs the mental dynamics by determining the answers to such questions as—what did this signal indicate to the problem solver about process state? Given a particular action, in what perceived process state or context is this action reasonable? Errors are seen as the result of limited rationality—the people involved in the incident are doing reasonable things, given their knowledge, their objectives, their point of view and limited resources, for example, time or workload (Reason & Mycielska, 1982; Woods & Roth, 1988). Reconstructing a trace of the problem-solving process can identify points where limited knowledge and processing led to actions that are clearly erroneous from hindsight.

In the end, any reconstruction is a fictional story—it may have happened "as if . . ." A critical factor is identifying and resolving all anomalies in a potential interpretation. We have more confidence in, or are more willing to pretend that, the story may in fact have some relation to reality if all currently known data about the sequence of events and background are coherently accounted for by the reconstruction. However, any reconstruction is tentative, for a later investigator may turn up more evidence that creates anomalies in previous reconstructions and motivates the creation of a new or modified account.

There are a number of major outstanding questions about how to do this type of critical incident study so that meaningful interpretable results are generated and not just anecdotes (cf. Klein, 1989b, on the alternative interpretations of the psychological implications of the Vinncennes incident). For example, for retrospective studies to be meaningful, should the investigative team personally interview participants in the incident and related domain personnel, or can the analysis be carried out based on second-hand reports? When? Are there ways to do retrospective analyses of decision making that support constructive debate about alternative interpretations rather than ad hoc assertions (e.g., when is it meaningful to conclude that an incident contains an example of people committing a decision bias or some category of human error)?

FIELD OBSERVATION

Another source of techniques for the study of human cognition outside the laboratory is the tradition of field studies in industrial settings (primarily European; see De Keyser, 1990, for an excellent critical review) and of anthropological field research (e.g., Suchman, 1987; Lave, 1988; Hutchins, 1980, 1983).

A field research perspective raises questions about the relationship of the investigator to the domain of interest and the domain practitioners. Do you have to "go native" or "become an expert yourself" in order to do meaningful complex world research (Hutchins, 1980)? Does the researcher require a domain-knowledgeable guide or informant (a Virgil to guide the researcher cum Dante through the seven circles of naturalistic research hell) to help penetrate the surface veil of the domain and identify the deeper structure (e.g., Cook & Woods, 1990)?

Meaningful investigations of complex behavioral situations where the domain practitioner's performance and skill is the focus of the study will require a significant amount of personal knowledge acquisition and experience with the domain, and especially with the role and point of view of the practitioners within the domain. Some of this domain appreciation can be provided by domain-knowledgeable guides; for example, the earlier discussion of behavioral protocol analysis mentioned several ways that other practitioners can be harnessed to help in data collection and analysis. Very frequently, it may be critical to formalize this knowledge acquisition through in depth cognitive task analysis (Roth & Woods, 1989).

While immersion in the domain "culture" is an important contributor to doing complex world studies, it is not the end itself. The danger is that one can be drawn in too deeply and learn only about that specific context. In part this is due to referential transparency—what one sees with is seldom what one sees (Hutchins, 1980). The investigator must preserve some distance (while at the same time being intimate with the domain details) in order to be able to see what the domain practitioner sees with.

The field research tradition points out a variety of techniques and obstacles in the study of complex behavioral settings (cf. Roberts & Rousseau, 1989). For example, one important requirement is to live among the "natives" long enough to be confident that you are minimizing the distortion produced by your presence prior to collecting any data. On the other hand, the practitioners in the field are not and cannot be treated as "subjects" in a traditional laboratory experiment (hence, the use of the moniker *study participant* in this chapter). Frequently, the reason an investigator has access is to provide practical assistance in solving "applied" problems. Providing this assistance is the coin of the realm for the time and real cooperation of the practitioner.

Is the Choice Between Rich Field Work and Spartan Experiments?

One possible methodological conclusion for those interested in complex behavioral situations is that there is no relationship to spartan experimental psychology laboratory methods. Rather, one should learn, use, and advance the techniques worked out for anthropological field research in order to do a "cognitive anthropology" of work cultures (Hutchins, 1980). I do not believe that these two approaches exhaust our repertoire. I am convinced that there are techniques for examining complex settings that fall between descriptive field observation and narrow laboratory experimentation.

Field observation is a valid, meaningful technique that belongs in our research repertoire. For example, field observation is necessary to establish the mapping between target and test situation, to make decisions about what manipulations to use in the test situation to make observable the phenomenon of interest while preserving the basic character of the target situation. Bartlett's (1932) investigations of cognition began with observation of everyday activities, which formed the basis for experimental studies, which in turn informed further observation in the field.

There are a variety of major problems in field research oriented toward understanding human cognition (cf. De Keyser, 1990). One that occurs even with good field studies is the gap between data collection and interpretation. The problem is that a critical reader of a field study report cannot retrace the study and reinterpret the purported conclusions, as one in principle can do with archival reports of traditional laboratory studies. This is a major issue for complex-world research. There tends to be a great leap from the data collected to interpretative conclusions, with a vast wasteland in between. This forces other researchers either to accept or reject the investigator's interpretation rather than criticize, reinterpret, or build on the study. This is exacerbated because it is not standard practice of authors or journals to include all protocols, at some level of encoding, in the report of a study which used a process-tracing method (but cf. Roth, Bennett, & Woods, 1987, for one exception). Effective methodologies for studying complex behavioral situations must support this process of criticism, reinterpretation, and follow-ons to produce cumulation, generalization, and the growth of knowledge.

ISSUES IN USING PROCESS-TRACING METHODS

The driving assumption behind this chapter is that there are research methodologies, which fall between the poles of descriptive field observation techniques, which can investigate complex behavioral settings directly, and spartan laboratory research approaches, which are relevant only obliquely to complex settings. In this section I will try to outline an approach to process-tracing studies that falls between these two poles.

The Concept-Specificity/Context-Independence Tradeoff

One technique (Hollnagel et al., 1981) to deal with the above gap in studies of complex behavioral situations is derived from the idea that there is a tradeoff between concept-specificity and context-independence in analyses of human behavior. The technique Hollnagel et al. proposed is to use a succession of levels of analysis in a process-tracing study, which begin at a context-dependent, concept-independent level of analysis. Performance is first analyzed or described in the language of the domain/profession: this user, in this domain, in this simulated or actual context, in this scenario, did action x or made statement y at time z. Analysis at this level—what can be called a *description of actual performance*—is relatively concept free but highly context dependent; in other words, it is difficult to generalize the results to other users, tasks, or events.

In the Hollnagel et al. technique the description of actual performance is followed by successive stages of progressively more concept-dependent, context-independent levels of analysis. The use of a non-domain-specific concept language based on models of human performance allows one to produce a description of behavior that is context independent and therefore potentially generalizable to similar situations in different domain contexts (cf. Montgomery & Svenson, 1989a). Since concepts or models form a basis for succeeding levels of analysis, they can be called *formal performance descriptions*.

Take an example from studies of human error (Reason & Mycielska, 1982). Imagine a user executing action set Y (an infrequently performed task) who erroneously substitutes actions from set X (a frequently performed and closely related task). The actual performance description would state that the user committed an error in maneuver Y, executing an action from set X, rather than the correct action from the appropriate action set Y. The formal performance description would state that a "capture" error had occurred, because the action in its domain context meets the criteria for a capture error defined as one category of human error. In this example, concepts about human error have been used to encode the domain level view of user performance; as a result, the data can be combined with, and generalized to, other users in other events. However, note that, despite the shift in language of description, the result can be seen as data (i.e., a description of what happened), albeit filtered through a conceptual looking glass. Furthermore, the conceptual dependence that underlies the abstraction can be specified explicitly.

In process-tracing studies of human problem solving and decision making the concepts used to move beyond context-dependent descriptions come from human information processing defined very broadly (cf. Pew et al., 1981; Woods, O'Brien, & Hanes, 1987; Roth et al., 1987; Klein, Calderwood, & MacGregor, 1989). In other words, the formal description of performance marks a shift from a domain specific language to some type of a cognitive language. This shift is often referred to as the encoding stage of analysis. It is important to keep in mind that the two descriptions exist as parallel accounts of behavior in the episode (trial).

"Field Experiments"

One can use the process-tracing approach described above to produce behavioral protocols that address cognitive activities in the incidents in question. But what is the larger context of the study? Protocol analysis or process-tracing is just another measurement technique (like reaction time measures). What defines studies in a larger sense is the psychological topic or concept being investigated. In more descriptive research, the psychological topic may be the kind of cognitive situation selected or staged for investigation. For example, how do people solve garden path problems: where this class of problems is defined (i.e., problems where there is a highly plausible but in fact erroneous hypothesis), a set of problems that have these defining characteristics is identified or created, and results from this investigation can be compared critically to other studies on this cognitive topic, independent of the domain that generates the specific problem (e.g., Johnson et al., 1981).

Let us pick up again the capture error example discussed a little earlier. The details behind the concept of a capture error provide guidance about what kind of situation should be created in order to have the opportunity to observe capture errors or investigate underlying psychological mechanisms. The concept specifies the critical variables, or the effective stimuli, to be measured or controlled—that is, what are the aspects of the situation that really matter with respect to the behavior of interest. As a result, the domain description of an episode can be shifted to a cognitive description in terms of the concept of capture error and similar error forms (slips). For example, distractions and interruptions may be critical contributors to the occurrence of slip errors. Therefore, the test scenarios should include these elements (note that an *interruption* must be defined as a domain-specific event, for example, a call from air traffic control timed to occur during the execution of a checklist, if the domain is commercial aviation flightdecks). Also, a formal description of behavior can be developed in terms such as the called-for action sequence, the timing or form of the interruption, the action sequence that could take over (the potential capture sequence), the participant's behavior following the interruption (repeating actions, ommitting actions, reorienting behaviors, etc.), and the relation between the nominal and the capture action sequences (e.g., frequency of occurrence, task criticality, etc.). In this approach, the traditional problems of identifying evaluation criteria and selecting test scenarios are mitigated, because the explicit formulation of the detailed psychological question to be tested contains a measure of successful performance and the essential conditions that must be produced in any test scenario.

What is fundamental in a protocol analysis study is the psychological question under study. This question guides the construction of test scenarios, the raw data to be gathered, and the kind of concepts brought to bear to analyze human behavior. Therefore, one can think of studies designed in this approach as field experiments—*field* experiments, based on the use of complex behavioral situations (which could be the naturally occurring situation or practitioner behavior in simulations at various levels of fidelity); field *experiments,* in that the scenario, study participants and conditions of observation are focused in on some psychological question.[3] Hogarth (1986, p. 445) has remarked that the yield of this type of study "depends crucially upon whether the investigator adopts an 'experimental framework' in organizing observations." Adopting this experimental framework means conducting the study as an empirical confrontation, that is, a process of observation where doubts can be formulable in order to reappraise beliefs.

Because of pragmatic limitations on *access* when one wishes to study actual practitioners working with substantive tools and on substantive problems, coupled with challenges of experimental design, field experimentation involves an element of capitalizing on naturally occurring opportunities to pursue a cognitive issue. Upon recognizing that the opportunity for access is available under conditions that will allow some control/focusing, the investigators proceed to shape that naturally occurring situation into an experimental investigation. Roth et al. (1987) is an example of this occurring. Pragmatic and scale limitations may preclude individual studies from including a thorough sample over conditions (Hammond, 1986b), but the explicit mapping between the test situation being investigated and the psychological situations and issues of interest allows for a cumulation of results and knowledge across studies. This is critical in order to avoid an endless stream of studies with apparently conflicting or nongeneralizable results (deKeyser, 1990).

Finally, the idea of studying complex settings directly, through tech-

[3] I have adopted the term suggested by Jane Malin of NASA Johnson Space Center—*field experiments* or *experiments in the field*—to describe this type of focused field-oriented study. Others have suggested other terms; for example, Gentner and Stevens (1983, p. 2) used "designed field observation, in which an artificial domain is constructed that has interesting relevance to the real domain under consideration."

niques like field experiments, reveals a hidden bias in both the "basic" and "applied" psychological communities. It is accepted without quarrel that studies of spartan situations will eventually result in fundamental concepts, laws, and models of cognitive processing that will transfer eventually to real world, applied problems, at least in principle. Meanwhile, applied researchers, driven by pressing problems related to people in complex systems, undertake domain specific applied research, frequently with few ties to the spartan basic research going on in parallel (hence, we have the field of human factors divided up by domain boundaries—aviation human factors, nuclear power human factors, forensic human factors, etc.). However, there is another possibility, one that is not accepted as a viable approach by either the basic or applied communities. One can study complex worlds directly and produce results that add to the generic research base on human cognition, as well as produce results specific to the complex setting that served as the research vehicle. The results will cumulate and be transportable from one complex setting to another with similar "deep structure," or even to the next problem in the same domain. One can call this approach a complementarity assumption on the relation of "basic" and "applied" behavioral science research.

The complementarity assumption maintains that complex settings have a dual interpretation: one as an "applied" or local problem to be solved within some temporal and resource horizon, and another as a specific behavioral context that is an instance of some classes of behavior which can serve as a field laboratory for investigating that class of behavior. As a result, there is a natural complementarity between growing the research base and using the research base to develop pragmatic, though approximate, solutions to application problems.

One criterion for understanding of a phenomenon requires that the scientist demonstrate control of the phenomenon, that is, the ability to eliminate or create the phenomenon, to enlarge or diminish the effect. If we claim to understand decision making, planning, problem solving, etc., then we must be able to demonstrate control of these aspects of behavior even in complex "applied" settings. In other words, we must be able to improve human performance through a variety of means for supporting human problem solving and decision making. Thus, demonstrating fundamental understanding can, at the same time, help solve immediate problems such as the impact of new information technology on the flightdeck on aviation safety.

Similarly, the only path to get ahead of the pace of technology change and the progression of domain-specific "hot buttons" is to use a generic but relevant research base to go beyond technology-specific or completely domain-specific descriptions (e.g., Cook & Woods, 1990).

This research base has been lacking or remained impoverished because of the gulf between basic and applied research. As a result, human factors and person–machine system researchers have been reduced to chasing a rapidly changing series of locally defined and technology-driven problems.

Field Experimentation Using Process Tracing

This section outlines a basic set of steps to be followed in setting up a field experiment using the process-tracing methodology.

Step 1. The critical precursor for all of the steps is to define the psychological issue being studied.

This can be done in several directions, that is, starting with an issue and then searching for an accessible situation that is an instance of the class to be investigated, or starting with an accessible complex setting and defining the class of psychological concepts or models that are relevant to that situation. Field observation frequently is an important activity in this step.

One typical error in the design of field experiments is to mistake superficial labels (e.g., diagnosis, planning) for insight into psychological issues. The source of this flaw is a failure to build an adequate account of the task demands—a cognitive task or work analysis or a cognitive model based in part on field studies.

Step 2. Develop an explicit mapping between the psychological issue under study and the test situation, for example, how does the question under investigation (e.g., garden path problems or the disturbance management cognitive task) get translated into characteristics of a specific test situation. Note that the test situation has a dual interpretation—one, in terms of domain-specific features and events, and a second in terms of a cognitive or behavioral language, that is, a behavioral situation that possesses certain characteristics.

The design of the study as an experiment occurs at this stage. The experimenter takes steps to ensure that the data of interest will be in the protocols to be collected. To accomplish this, it also helps to define the protocol building process, that is, the raw data collection step, so that the investigators will be in a position to extract the data of interest from the raw protocols. The experimenter develops the study as a field experiment, primarily through manipulation of the set of scenarios and the detailed features of individual scenarios in relation to the psychological questions of interest. For example, Roth et al. (1987) developed a set of scenarios where the goal was to challenge straightforward fault diagnosis in order to learn about some issues in human–

intelligent computer cooperation. As a result, the problem set was selected to include cases with various kinds of complicating factors such as miscommunications, impasses, bugs in the knowledge base, and multiple faults.

At this stage the experimenter should develop the tactics to cope with the challenges to validity in process-tracing studies—tactics for generating observable data that will support inferences about the psychological topic of interest. How do the assessment techniques change the primary processes that are the target of study, and how can I eliminate, counterbalance, or estimate the effects of these changes? How can I minimize or place checks on omission of important aspects, intrusions of irrelevant features, or distortions of the actual processes in the data collection and analysis process? Consider a simple example of the tradeoffs the experimenter faces. In dynamic fault management situations the portions of a scenario that are of greatest interest are almost always high operator demand periods where any secondary task can interfere or interact with the primary task. Often in studying this type of situation the experimenter wants to understand how the problem solver copes with high workload. A thinking-aloud technique is not suited to these circumstances, as it constitutes a secondary task that may interfere with the phenomenon under study or that may be shed at exactly the time when the experimenter wants to focus his or her data collection efforts.

Step 2 is also important in avoiding the psychologist's fallacy. As mentioned earlier, one is not justified in *assuming* that the experimenter's representation of the problem is the same as the study participant's representation; that is, the problem representation is not inherent or given in the problem statement *outside any larger context*. The experimenter has the burden to include some manipulation or check that provides evidence about the participant's problem representation. For example, Maule (1989) combined a classic laboratory technique with protocol analysis to investigate the role of decision frames in choice problems. One of the advantages of process-tracing methodology is that the investigator directly focuses on determining the participant's representation of the problem.

A typical danger at this stage (besides failing to build any mapping between test and target behavioral situations) occurs in studies that use an experimenter-designed microworld as the test situation. The microworld may be related to the target situation, but only at a surface level, for example, cloaking the microworld in the language of the target situation, which masks the absence of any deeper relationship based on a cognitive model or task analysis (cf. Cook & Woods, 1990).

Step 3. Collect data; that is, run the study participants in the test scenarios.

Step 4. Construct a domain specific protocol from the raw data records for each episode (trial).

Process-tracing studies generate a very large amount of raw data. One frequent source of failure in studies of complex worlds is getting lost in the details of overwhelming amounts of raw data. A critical pragmatic issue is efficiency in processing of the raw data. Using the knowledge of the issues being investigated, and the encoding approach to be used in step 5 to focus the collection and handling of raw data, can greatly increase the efficiency of the process (Woods, O'Brien, & Hanes, 1987). However, note that, in this filtering process, one should not lose the base raw data, in part because interesting behaviors and findings that were not anticipated in detail in advance almost always occur in studies of complex settings.

Step 5. Construct a formal, cognitive, or psychological protocol for each episode.

The cognitive encoding of the description of actual performance is the critical step in producing general, transportable results. Note that, frequently, there should be several layers of these protocols. One can see this first layer as a translation or transformation of the raw domain data into a cognitive language. Successive levels of analysis can attempt to get more leverage through stronger conceptual dependencies. In this process it is critical to be able to separate uncertainties regarding what occurred from debates on interpretations about what the observed behavior means relative to some psychological issue.

Another issue in human performance studies that use process-tracing concerns the reporting of data. Process-tracing studies, like other studies, should be reported in a way that supports a critical reading. One limit on complex world studies is that other researchers need to understand the domain in detail in order to be able to examine the study critically. Obviously, reporting lengthy process tracings also presents difficulties. The cognitive description, or at least a schematized version that minimizes the need for extensive domain background knowledge, should be required for archival publication. See the protocols as published in Roth et al. (1987) for an example of the entire technique described here, especially as an example of cognitive, domain independent protocols.

A typical failure in studies of complex worlds is to get lost in the language of the domain, that is, to fail to move from a domain description of behavior to a psychological description. This is due often to a failure to specify the mapping between the test situation and the target situation of interest.

Another difficulty comes from the danger of excessive microencodings of participant information processing. The fundamental point of a protocol analysis is to specify the process an individual (or a team) used to solve a particular problem, for example, to extract their strategies. The investigator's first responsibility is to be able to report these strategies. These strategies are the critical unit of analysis on which other aggregations and analyses of data are based.

Finally, there is always the danger of the hindsight bias. As was mentioned earlier, the point of a process-tracing method is to establish how the incident unfolded from the point of view of the person in the problem.

Step 6. Analysis across protocols with respect to psychological questions of interest.

One way to assess the protocols is to generate measures of the problem-solving process by relating behavior against a background frame as a model of the task.

This can be done in a variety of ways. One is to build a problem space for each test scenario that captures the set of trajectories along which the incident may evolve from the point of view of the person in the situation, at each point in the evolving incident. This includes mapping the cues potentially available, their interpretation in context, the knowledge relevant to characterize the underlying problem state, the potentially relevant goals, and the set of plausible actions. This is built through a cognitive task analysis of the domain. For efficiency purposes this should be done prior to data collection and used during Step 3 as an aid to generate efficiently both the raw and the first-level encoded protocols.

Another is building a cognitive simulation as an explicit model to account for the observed behavior (Newell & Simon, 1972). The concepts at the formal level of description can be as strong as a cognitive model for the class of situations investigated. Expressing this model as a runnable symbolic-processing simulation allows one to test the ability of the concepts captured in the simulation to account for the observed behavior. When the cognitive simulation produces the same behavior as the human study participants, the model becomes a candidate hypothesis about the underlying cognitive activities or strategies people use in that situation (e.g., Woods, Roth, & Pople, 1987). Furthermore, the formalization required to produce a cognitive simulation introduces increased rigor into the interpretation of the protocol results.

In effect, the problem space and cognitive simulation techniques are ways that one uses psychological concepts and models as a language for the formal layers of analysis. In addition, the problem space and cognitive simulation techniques can be used as computer aids for protocol analysis (e.g., Kowalski & VanLehn, 1988; there is also work to develop concept-neutral computer aids that support more efficient encoding and manipulation of protocol data).

Another type of background frame for interpreting participant behavior comes from what other domain practitioners see as plausible activities. A basic defining characteristic of human information processing is that the immediate problem-solving context biases the problem solver. In naturally occurring problems, the context in which the incident occurs, and the way the incident evolves, activate certain kinds of knowledge as relevant to the evolving situation, which affects how new incoming information is interpreted. The investigator can gain insight into the observed behavior by comparing it to how other domain practitioners would interpret the situation. In one technique the investigator presents the evidence available at one point in the evolving incident to observers who have not undergone the incident evolution. This "neutral" observer then makes a judgment or interpretation about the state of the world, relevant possible future trajectories, and relevant courses of action. The question is whether the path taken by the actual problem solver is one that is plausible to the neutral observers (i.e., they entertain that path as a serious candidate).

Special issues of interpretation of behavior arise when the focus of the study includes human error. The main problem is definitional—there are various positions about how to define errors including both domain standards and psychological taxonomies (cf., e.g., Reason, 1990).

Special issues also arise with respect to tools for problem solving. Spartan research related to problem solving strips the participants of any tools. Yet in almost all naturally occurring situations people devise, adapt, or utilize tools to assist in accomplishing goals or reducing effort. Understanding how people adapt or create tools may be an important approach to a better understanding of human problem solving. Studying human performance with significant tools adds another set of factors to an already multifaceted stimulus situation (cf. Hutchins, 1989, and Cook et al., 1990, for examples of such studies; cf. Woods, O'Brien, & Hanes, 1987, for some of the methodological challenges in studying tool use). What role does the tool play in practitioner cognitive activities (e.g., does it function as an external memory)? How are practitioner strategies changed when the tools available change? In this case the difficulty is that questions about the cognitive role of problem-solving support systems and tools are framed too easily in terms of the technologies from which the systems are built (e.g., should I use tiled or overlapping windows?). But again we are led to the challenge of developing a cognitive language of description, not only independent of the domain language (i.e., the language of the complex setting itself), but also independent of the languages of tool creation (Woods & Roth, 1988).

Another challenge for process-tracing methods is describing activities involving distributed cognition (Hutchins, 1989), where the cognitive activities in monitoring, situation assessment, and corrective actions are distributed over several people. With the introduction of intelligent machine agents into systems, cognitive activities become distributed over people and machines (Woods & Roth, 1988).

DISCOVERY OR VERIFICATION?

Destructive debates arise from claims that one methodological strategy or another has priviledged access to fundamental results. The contrast between spartan laboratory situations and complex behavioral settings was used as a vehicle to point out the underutilization of direct investigations of complex settings. Convergence between studies of simple and complex behavioral situations is important—there is a tradeoff, with sources of uncertainty and error on both sides (Bartlett, 1932; Hogarth, 1986).

One strategy may be more appropriate for hypothesis generation or discovery as opposed to hypothesis testing or verification, especially in immature research areas such as person–machine interaction. For example, studying complex settings can help to focus spartan lab research in more productive directions; spartan lab research results can guide new ways to look at and parse complex settings (Bartlett, 1932, saw this as the proper relationship between behavioral science research directed at spartan and complex settings). One research problem is revealing the basic phenomena, given the richness of the phenomenal world—seeing beyond the "blooming, buzzing confusion." Before we can say what hypothesis accounts for some observed effect, we need to characterize what are the observed effects, especially in terms that are not test situation, tool, or domain specific. We need to develop new ways of seeing in these rich situations—rich in the knowledge of the participants, rich in the diversity of strategies by which one can achieve satisfactory performance, rich in the tools available to assist human performance. In many areas the current need is to generate meaningful, fruitful hypotheses that eventually may be testable under more targeted circumstances.

In a theory-testing approach to behavioral science research, the objective is to support the current model and extend it to new tasks/variables. The critical constraint in developing the experimental task is tractability within the theoretical paradigm. The emphasis on the mapping between target and test behavioral situations points to another valid research strategy—one that is task driven where a cognitive analysis of the task demands guides the specification of meaningful test situations and relevant psychological models and concepts (e.g., Sorkin & Woods, 1985; Kleinmuntz & Thomas, 1987). The cognitive analysis then guides how the pieces (existing data models or new results) are put back together to form a coherent picture of human behavior in that class of situations. In the task-driven strategy, there is a natural complementarity between growing the research base and using the research base to develop pragmatic, though approximate, solutions to application problems.

SUMMARY

Research on problem solving in more complex situations, where significant tools are available to support the human and where experienced domain knowledgeable people are the appropriate study participants, requires a shift in research methodology from typical laboratory studies. This does not mean that rigor or control or generalizability or theory must be sacrificed. Rather it means using a wide set of research tools to converge on an understanding of the phenomenon in question.

> The analyst's task is no more difficult in the field setting than in the laboratory. The impression that this cannot be so rests primarily on unjustified assumptions regarding the extent to which the behavior of subjects in experimental settings is directly revealing of cognitive processes. Whether the setting is experimental or natural, the investigator must be able to make and support assertions about the representation of the task that the subject brings to the task. The laboratory setting has its advantages, but so has a more naturalistic environment. (E. Hutchins, 1980, p. 125)

KEY POINTS

- It is not necessary to impose a spartan research methodology to achieve tractability in studying complex behaviors.
- Whether data are collected from laboratory or field studies, the concern should be representativeness and the mapping between test situations where one is observing behavior and target situations which one wishes to understand.
- One methodology available to study complex behavioral situations is process tracing (or protocol analysis).
- There are numerous types of process-tracing methodologies that can be imposed on various field study designs to yield reliable and generalizable results.