HUMAN-COMPUTER INTERACTION, 1994, Volume 9, pp. 251–317

Exploratory Sequential Data Analysis: Foundations

Penelope M. Sanderson
University of Illinois at Urbana–Champaign

Carolanne Fisher
MAYA Design Group

ABSTRACT

Human–computer interaction (HCI) investigators must consider the sequential nature of interaction and must often weigh behavioral, cognitive, and social factors when studying and designing today's increasingly complex systems. In many cases, laboratory experimentation is inappropriate and formal modeling intractable; instead, observational data analysis is frequently the only appropriate empirical approach. Diverse approaches to observational data analysis already exist, which we synthesize as instances of *exploratory sequential data analysis* (*ESDA*). In this article, we outline fundamental ESDA characteristics that might help HCI investigators using sequential data make better conceptual and methodological choices. ESDA owes a philosophical debt to exploratory data analysis but focuses on exploring sequential data. Important issues for ESDA are finding an appropriate temporal band for analysis, finding an effective semantics for encoding, and completing an analysis in an acceptable time frame. We survey temporal factors

Penelope M. Sanderson is an engineering psychologist who, in her investigations of cognitive engineering and human–computer interaction, seeks to understand cognition in complex settings; she is an Associate Professor of Psychology and of Mechanical and Industrial Engineering at the University of Illinois at Urbana–Champaign. Carolanne Fisher is a cognitive scientist who specializes in designing high-technology products that are easy for people to use; she is Director of Human Sciences at MAYA Design Group.

CONTENTS

1. INTRODUCTION
2. EXPLORATORY SEQUENTIAL DATA ANALYSIS
    2.1. Definition and Overview
    2.2. Practical and Conceptual Constraints on ESDA
3. THE TEMPORAL WORLD OF ESDA
    3.1. Spectrum of Sequential Events Relevant to HCI
    3.2. AT:ST Ratio and Movements Across Spectrum
4. EXPLORATION AND THE EIGHT Cs
    4.1. The Legacy of EDA
    4.2. ESDA Smoothing Operations: The Eight Cs
5. ESDA TRADITIONS AND TECHNIQUES
    5.1. The Behavioral Tradition
        Typical Characteristics
        Example: Hypothesis Testing With Observational Data
        HCI and the Behavioral Tradition of ESDA
        Boundaries of the Behavioral Tradition
    5.2. The Cognitive Tradition
        Typical Characteristics
        Example: Cognitive Process Modeling
        HCI and the Cognitive Tradition of ESDA
        Boundaries of the Cognitive Tradition
    5.3. The Social Tradition
        Typical Characteristics
        Example: Interaction Analysis
        HCI and the Social Tradition of ESDA
        Boundaries of the Social Tradition
6. ESDA PROBLEMS AND ADAPTATIONS
    6.1. Failures of Expertise
    6.2. Time Management: AT:ST Threats
        AT:ST Ratios and ESDA Traditions
        Benefits and Traps of Software and Hardware Support
    6.3. Data-Based Problems
        Promoting Adaptation by Composing the Eight Cs
        Qualitative Impact of ESDA Software
    6.4. Challenges of Video and Reporting ESDA Studies
7. CONCLUSIONS

and introduce *analysis time:sequence time ratios,* which describe the time cost of conducting different types of ESDA. We also introduce the "Eight Cs"–different general transformations that can be performed on sequential data. We conjecture that the Eight Cs, and their combinations, are critical for supporting scientific inference in ESDA. Distinctions are made among three principal ESDA traditions that are relevant for HCI–behavioral, cognitive, and social. We indicate how each

ESDA tradition has been used in HCI and describe one technique from each tradition. Last, we outline major practical problems for investigators using observational data and, following our framework, suggest ways such problems might be overcome.

1. INTRODUCTION

Most software is designed to sustain episodes in which individuals or teams achieve some result interactively over time with the help of computing and/or communications equipment. In typical episodes, we might format manuscripts, compose and mail electronic messages, explore and analyze scientific data, attend computer-supported collaborative design meetings, safely land airplanes, or hold videoconferences with remotely located colleagues. Such performances are sequential in nature. Thus, if we are to understand them, design for them, and evaluate their success, then very often we need to make observations of how they unfold. Moreover, such performances typically have behavioral, cognitive, and social characteristics. So, given records of human–system interaction, we can note actions by human agents, gather evidence for human thought processes, and see how humans interact with both engineered artifacts and other humans.

For many human–computer interaction (HCI) researchers and practitioners using observational techniques, it is increasingly necessary to view HCI as a rich multidimensional activity that unfolds over time–one in which data exploration is not only a virtue but also a necessity. However, with modern recording technology, observational data are much easier to collect than they are to analyze, and analysis is often prohibitively time-consuming. Even given the time, we are less and less sure what type of data we should be accessing and what form our analysis should take. What techniques can be used?

Many analysis techniques that use naturally unfolding events as data have been used in the service of HCI research and design. These include protocol analysis, video analysis, discourse analysis, conversation analysis, interaction analysis, cognitive task analysis, and sequential data analysis. A distinct blend of intellectual traditions lies behind each technique. Unfortunately, however, research and design questions do not always unambiguously suggest one technique, and the nature of the data we collect is starting to breach the traditional boundaries of these techniques. As G. M. Olson, J. S. Olson, and Kraut (1992) put it:

> The behavioral science side has extended to the more social disciplines such as social psychology, organizational studies, anthropology, and sociology. There has been rapid diversification in the empirical methods used to study

> people working with new technologies, with increased reliance on observational methods (often exploiting video recording technology) and inclusion of such approaches as ethnomethodology and conversation analysis. (p. 252)

The implication for HCI research is that it is desirable and often necessary to view a research question from multiple perspectives. Observable activity, inferred cognitive processes, and social interactions may each by itself be insufficient for a full understanding of how humans behave in complex computer environments. Qualitatively different types of data have to be integrated and understood. Researchers need to go beyond their own disciplinary biases or training, expanding the range of questions they are prepared to tackle and broadening the research techniques they are prepared to adopt.

However, choosing between existing techniques and finding new blends poses a challenge. It is not always clear which observational technique–or which combination of techniques–will be most appropriate for a given research question. To make an informed decision, the researcher must understand each technique very well–its intellectual tradition, the questions it was developed to answer, and the type and treatment of data it requires. Only then can the coherence of new combinations of observational techniques be assessed. Good research has always involved adapting methods and tools to new uses, but adaptations always require a thorough understanding of the trade-offs involved.

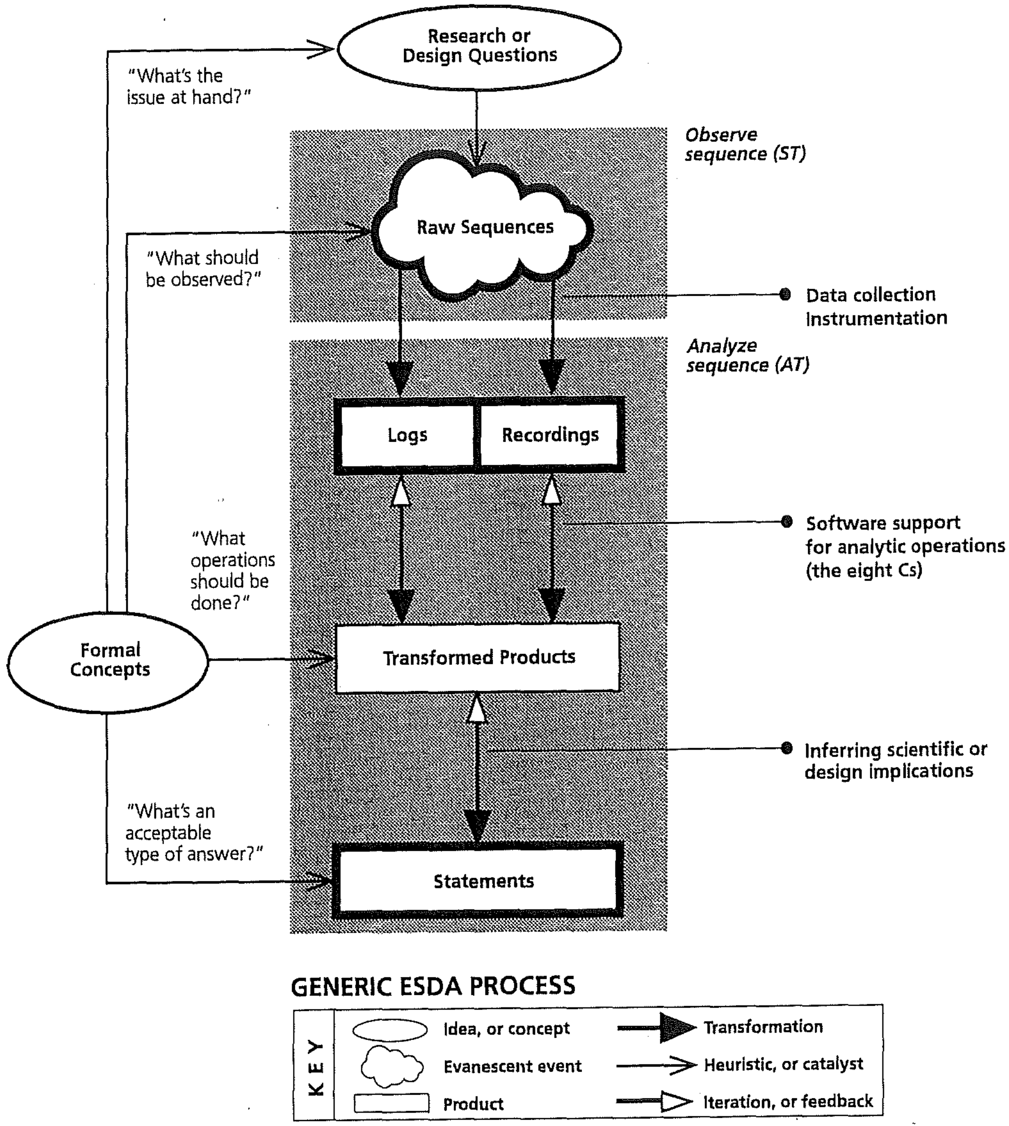
The goals of this article are to identify the constraints and traditions underlying current approaches to observational data in HCI and to indicate how researchers might synthesize powerful yet manageable new approaches to such data. We (a) define the scope of observational data and draw a parallel between the exploration of observational data and other forms of exploratory data analysis; (b) examine generic operations that support a researcher's ability to draw conclusions from observational data; (c) introduce the behavioral, cognitive, and social traditions of observational data analysis and provide detail about one specific technique from each tradition, noting relevant software; and (d) identify problems that researchers commonly encounter when working with observational data and map ways they might adapt techniques to best suit their needs.

2. EXPLORATORY SEQUENTIAL DATA ANALYSIS

2.1. Definition and Overview

Any kind of observation involves examining certain activities as they unfold sequentially over time. *Exploratory sequential data analysis* (*ESDA*) is the term we have chosen to describe the use of such data in the human sciences (Fisher & Sanderson, 1993; Sanderson, 1991). We define it as follows:

Figure 1. **The generic ESDA process.**



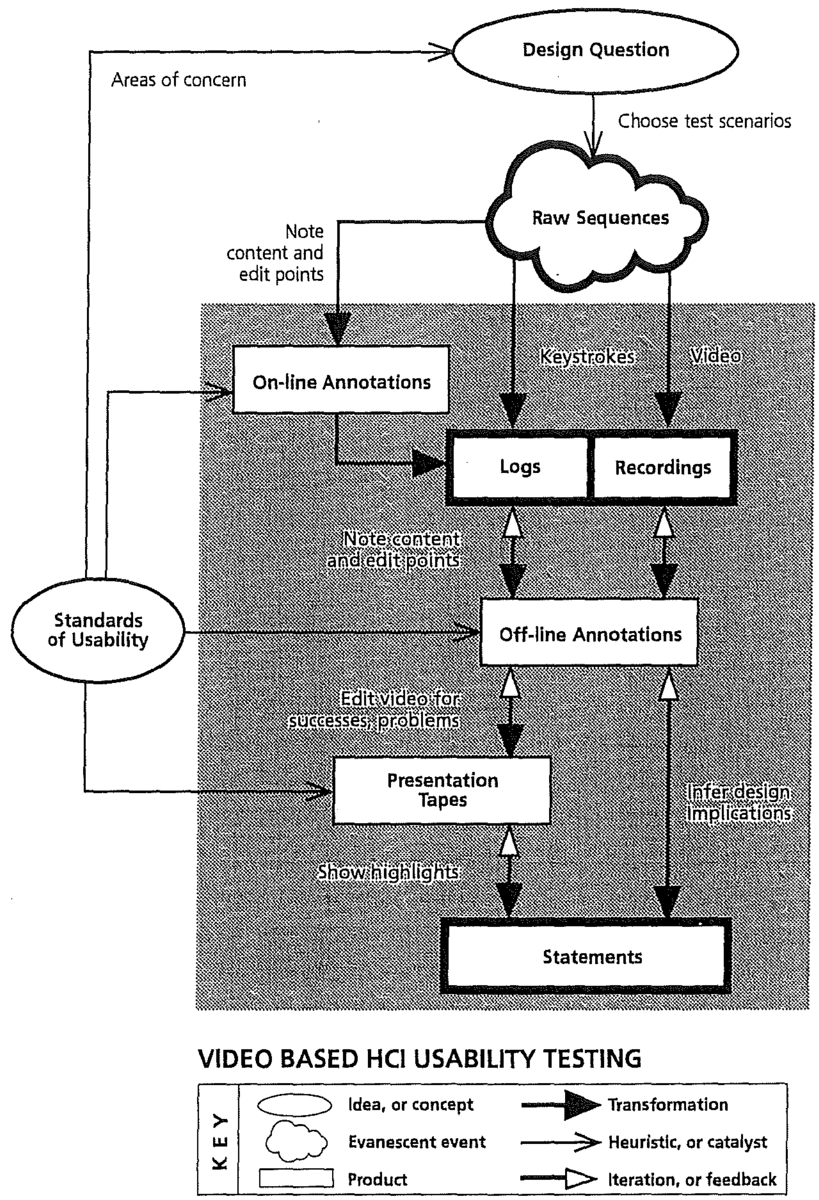
ESDA is any empirical undertaking seeking to analyze systems, environmental, and/or behavioral data (usually recorded) in which the sequential integrity of events has been preserved. The analysis of such data (a) represents a quest for their meaning in relation to some research or design question, (b) is guided methodologically by one or more traditions of practice, and (c) is approached (at least at the outset) in an exploratory mode.

Figure 1 describes in the most general way how ESDA is performed. ESDA involves moving from some research or design question to a final statement under the guidance of certain formal concepts and analytic practices. Raw Sequences, Logs and Recordings, and Statements are events or products present in any type of ESDA. *Raw Sequences* are the fleeting system, environmental, and behavioral events that are observed. *Logs and Recordings* are the aspects of these events that are captured either as computer data logs or on videotape or audiotape. Logs and recordings must be integrated for analysis. *Statements* are the conclusions resulting from ESDA that answer the initial research or design questions in a way that is meaningful to others. Two other nodes represent ideas and concepts that influence how ESDA is performed on a particular occasion: *Research or Design Questions* refers to the investigator's initial motivations for making observations, and *Formal Concepts* refers to the particular theoretical tradition and analytic techniques that an investigator may prefer. Last, the node *Transformed Products* represents the results of operations that investigators perform on data as they move toward their final statements. How the transformations are performed and what the products are like will be determined by the ESDA technique used and whether the purpose of the analysis is basic scientific research or design and evaluation. However, we argue that there is a basic set of analytic operations from which investigators using ESDA can draw in different ways (see Section 4.2).

The arrows in Figure 1 distinguish two relations between nodes. Arrows with solid heads indicate that data have been transformed in some way, whereas arrows with forked heads indicate the guidance of an idea or concept. Although the arrows linking the various nodes between Research or Design Questions and Statements might suggest that ESDA is a linear process, this is absolutely not the case. Many ESDA practitioners find that ESDA involves exploration, feedback, and iteration. One is often reviewing and changing "more transformed" ESDA products in light of "less transformed" products. For example, an investigator may wish to check coding or summary statements by reviewing the original data logs and recordings. Figure 1 indicates just some of these possibilities with arrows with empty heads that run in the opposite direction from the arrows with solid heads. Such exploratory activities are an important part of ESDA, and the term *ESDA* is deliberately redolent of Tukey's (1977) exploratory data analysis (EDA; see Section 4.1).

As Figure 1 indicates, Formal Concepts have a pervasive influence on the ESDA process. A certain outlook, such as cognitive versus situated action, will influence the way a research or design question is framed ("What's the issue at hand?"), the data deemed relevant for answering the question ("What should be observed?"), the transformations that will bring the investigator closer to an answer ("What operations should be done?"), and what an investigator's audience will consider a valid and well-substantiated answer ("What's an acceptable type of answer?"). We outline the influence of formal concepts when we discuss how the behavioral, cognitive, and social traditions of ESDA have shaped ESDA techniques. However, no exposition of the formal concepts underlying an ESDA technique and no description of it in practice will allow a newcomer immediately to

Figure 2. **Usability testing as an ESDA process.**



apply it effectively. In order to transform Logs and Recordings into Transformed Products and eventually into Statements that answer some question, investigators often must reinterpret their chosen technique, adapting it to the circumstances. A considerable amount of technical competence, formal background knowledge, and prior experience with the technique is required to adjust familiar techniques for novel circumstances.

To show how the generic ESDA process diagram provides a foundation for describing ESDA techniques, Figure 2 represents video-based HCI usability testing as an ESDA technique. Raw Sequences, Logs and Recordings, and Statements are present, as always. Formal concepts become the investigator's *Standards of Usability* for software, which may or may not be rigorously operationalized. There also may be a specific Design Question motivating the choice of scenarios to observe. Transformed products are few, to save time, but might include *On-Line Annotations* (notes taken during observation and edit marks on the video), *Off-Line Annotations* (notes and edit points made during data reviews), and, finally, *Presentation Tapes* (summary videotapes of usability problems or successes suitable for presentation to colleagues).

2.2. Practical and Conceptual Constraints on ESDA

Originally, observational data were collected manually by teams of trained observers who spent considerable time preparing for observational sessions. Many researchers trained in this approach viewed the introduction of mechanical recordings as a false and potentially dangerous luxury, promoting lax preparation and the accumulation of masses of unanalyzed recordings (Bales, 1950; Sweeney, Maguire, & Shackel, 1993; Woods & Sarter, 1993). Around the time that recording technology appeared, however, research questions started to emerge that depended heavily on review and post hoc analysis (Kendon, 1982). For example, the analysis of interpersonal interaction required repeated viewing of interactional sequences, and "the use of note-taking or on-the-spot coding of behavior cannot allow the researcher to recover the details of original materials and is indeed already analysis in itself" (Heritage, 1988, p. 131). In such research, on-line analysis is deemed inappropriate and premature. The research outcome instead depends heavily on a researcher's ability to work effectively through large amounts of uncoded raw data.

For HCI, as for any field, researchers must first ask whether ESDA is really the best research strategy for their question. The absolute amount of time required may be prohibitive, and many research questions can be answered with simpler and quicker methods. ESDA should be undertaken only when:

- The data needed to answer a question cannot be gathered any other way–for example, when the sequencing and temporal organization of activity are central to a question.
- The analyst has identified or adapted a viable technique for moving from Questions to Statements in the context of available Logs and Recordings and is competent to implement the technique.
- It is clear that the practical or conceptual benefit of using an ESDA technique is worth the extra time involved and that time is available.

The shading behind Raw Sequences in Figure 1 indicates the duration of an observed sequence (sequence time, or ST). The shading behind the lower nodes indicates factors that contribute to the time needed to analyze a recorded sequence (analysis time, or AT). The ratio of AT to ST (*AT:ST ratio*) usually lies between 5:1 and 100:1, but more extreme ratios are possible. An AT:ST ratio of 1:1 might be seen in the simplest type of observation-based usability testing, in which impressions gathered during on-line observation are refined in a single replay of the video recording, and design implications are immediately stated (Hammontree, cited in Weiler, 1993; Mayhew, 1990). Nielsen (1993) estimated the ratio to be between 3:1 and 10:1 for the analysis of video-based usability data, and Sweeney et al. (1993) provided a relatively conservative estimate of 10:1. In contrast, ratios of 500:1 or even 5000:1 (Ritter & Larkin, 1994) are seen in studies aiming to model the cognitive processes of users, in which fine-grained coding, analysis, model building, and testing are needed. Clearly, large AT:ST ratios impose an important practical constraint on a researcher's ability to perform ESDA effectively.

To reduce the AT:ST ratio, many researchers have started to build software to support ESDA, which we note throughout this article. These efforts are critical to the development of ESDA methodologies. In fact, we see ESDA as a topic for HCI as well as a methodology to serve it, because well-designed interactive software should help researchers perform ESDA. With complex sequential data, if there are no reasonably appropriate software tools to aid analysis, then methodological and conceptual development is hampered. If there is no methodological and conceptual clarity, however, then useful software tools are unlikely to be developed. In this article, we identify needs and advances on both fronts.

3. THE TEMPORAL WORLD OF ESDA

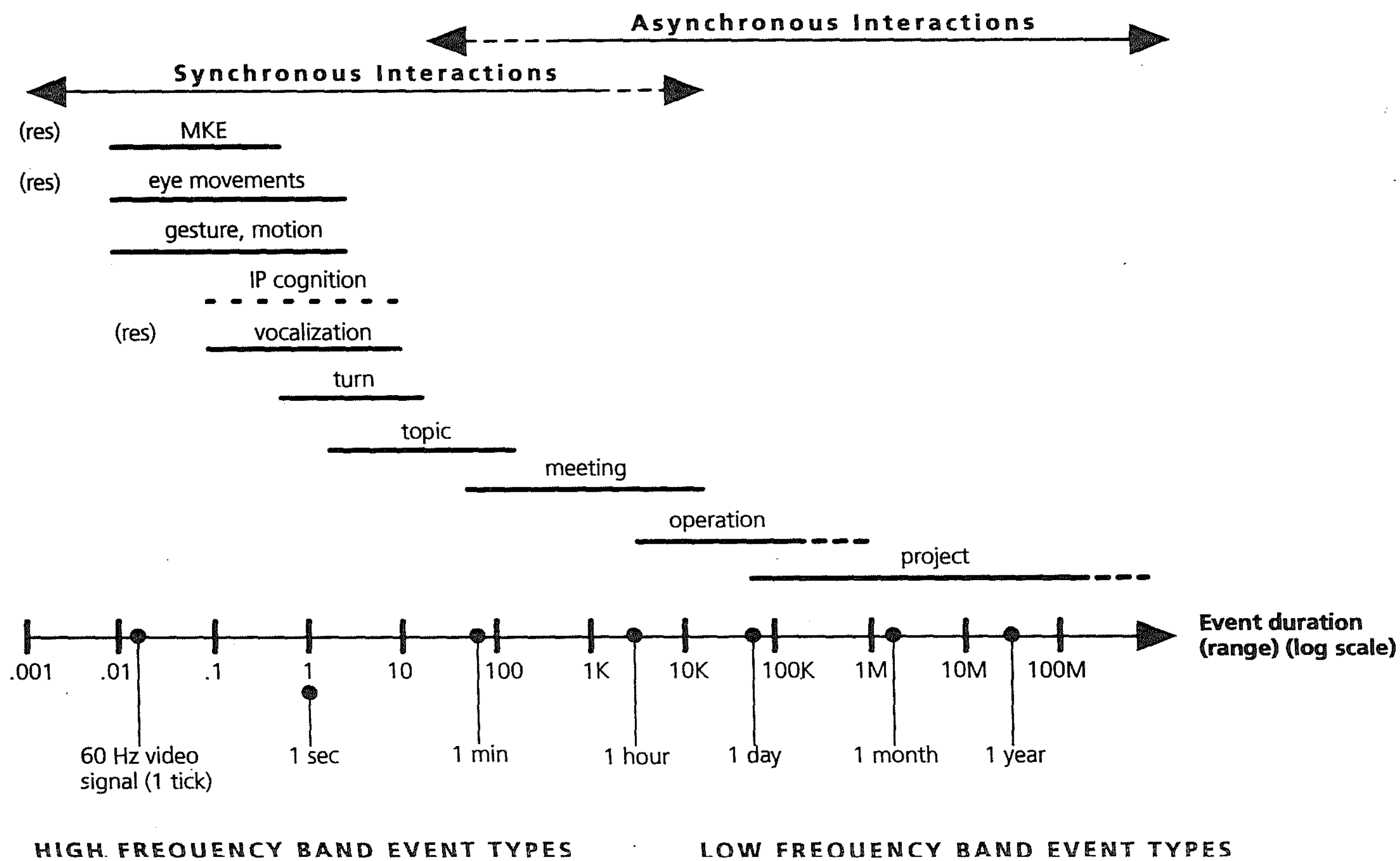
When performing ESDA in the field of HCI, many questions about time arise. First, what are the different kinds of sequential data we might deal with, and what is the time investment required? Second, what kinds of exploratory transformations and operations can help us extract answers to research questions from sequential data? Third, what are some constraints on the transformations and operations we might use?

3.1. Spectrum of Sequential Events Relevant to HCI

Before discussing different ESDA techniques and identifying fundamental ESDA operations, we must examine the spectrum of events that are typically collected in the types of observational studies performed in HCI and identify some of the constraints they impose on any kind of ESDA technique.

Figure 3 shows a spectrum of durations for different types of events that are important for research and design activities in HCI and that are typically captured in observational records. The horizontal log scale ex-

Figure 3. Spectrum of durations for different event types typically used in HCI studies. Approximate ranges for durations of individual events in HCI studies and approximate ranges for durations of individual events of each type are shown. In an analogy with the continuous domain, when sequences of discrete events of a certain type are considered, the typical number of events per second is the frequency of that event type. Event types ranging over short durations can show higher frequencies when in sequence (high-frequency band event types); event types ranging over long durations can show lower frequencies when in sequence (low-frequency band event types). MKE = mouse and keyboard events. Dashed line for IP (inferred processes) cognition indicates that these events are inferred rather than directly observed.



tends from time in seconds to years. The range of possible durations for individual events of each type extends between one and two orders of magnitude. Moreover, the ranges for different event types overlap. In an analogy with the continuous domain (e.g., electromagnetic energy or analog signals), the typical number of events of a certain type per second can be called the *frequency* of that event type. The maximum possible frequency for an event type, assuming that events arrive serially, is constrained by the duration of the events. For a given event type, events occur within a range or band of frequencies that, again by analogy with the continuous domain, we call a *frequency band.* Thus, event types ranging over short durations can show higher frequencies when in sequence and are referred to as *high-frequency band event types* (see left of Figure 3); event types ranging over longer durations can show lower frequencies when in sequence and are referred to as *low-frequency band event types* (see right of Figure 3). Research on the details of interface design has often focused on high-frequency band event types, whereas research on computer-supported cooperative work (CSCW) has focused more on mid- and low-frequency band event types.

As Figure 3 shows, mouse and keyboard events (MKEs), eye movements, and gestures range anywhere from 10 msec to about 1 sec. Cognitive processes (shown dotted to indicate that they are inferred) range from about 100 msec to 10 sec. Units of vocalization and conversational turns sometimes stretch to more than 10 sec. The smallest topic units (e.g., themes, speech acts) may be as short as 2 sec or as long as several minutes. Meetings range from a few minutes (imagine a quick conference call with two colleagues) to several hours. The range for operations (air traffic control, power generation) is harder to identify because operations often never close down, but they might be viewed as lasting from fractions of an hour to the length of a shift and beyond. Last, projects (collaborative design, software development) extend anywhere from 1 day to more than 1 year. Some properties emerge from this characterization that help clarify the challenges ESDA poses.

1. Sequences composed of higher frequency band event types usually occur synchronously. For example, a series of gestures, turns, or topics is generally seen in the context of a continuously recorded session. However, sequences composed of lower frequency band event types, such as meetings, operations, and projects, may occur asynchronously through, for example, messaging systems, computer-aided design software, memos, and letters. Different strategies for data sampling and capture are needed for these two cases.

2. Properties may be distorted by recording equipment. The markings "(res)" in Figure 3 indicate typical technical resolutions for finding the onset and offset times of high-frequency band events, assuming appropri-

ate instrumentation. However, if video is used to capture very fleeting events, then, because of the scan rate (30 frames per second for the NTSC video standard), some events may be missed, or there may be error in the time onsets and offsets stored.

3. Events within a given frequency band are often made up of events of a different type at higher frequencies. They also combine to form events of a different type at lower frequencies. For example, a conversational turn is composed of vocalization, gesture, and eye movements, and a series of conversational turns may combine to form a topic under discussion within a meeting.

4. Low-frequency band events do not directly reveal their composition from events at higher frequencies and so serve as "low-bandpass filters" (allowing only low-frequency information to pass). Taken individually, high-frequency band events do not reveal how they will combine to form events at lower frequencies and so serve as "high-bandpass filters."

3.2. AT:ST Ratio and Movements Across Spectrum

We can now return to the universal complaint that research involving observation is time-consuming. The time to analyze a sequence (an ordered set of events) usually is much longer than the real-time length of the sequence, assuming the latter has been captured synchronously, and the AT:ST ratio can vary a great deal. What determines the size of the AT:ST ratio? Imagine that we have a particular observational sequence of a given length, composed of particular logs and recordings, so that ST is a constant. AT will be affected by factors such as

- Data preprocessing or preparation time, such as the transcription of verbalizations or the preliminary parsing of action sequences to a useful frequency band.
- How easy it is to determine the event type and frequency band that will speak most clearly to each of the questions being posed.
- The frequency band of annotation or encoding required to test each question, determining the grain of analysis needed.
- "Observability," or the distance between what is literally seen and heard and what its significance is for the theoretical framework and question at hand. Is it easy to recognize the events or features of interest, or is considerable inference required to identify them and appreciate their meaning?
- The structural complexity (interconnectedness of concepts) of the analytic thinking required to develop a theory, test a model, or judge a design.
- The number of questions being asked and how well defined they are.

Clearly, if the theory or question being tested is complex and involves a detailed network of inferences, then the AT:ST ratio will be high. Simple arithmetic estimations suggest constraints on the quantity and quality of data analysis that will ensue. If a researcher found 3 unbroken hours a day for data analysis and worked linearly through the data at a given AT:ST ratio—we call this *rule-based linear coding*—then, at an AT:ST ratio of 5:1, the researcher would analyze 36 min of ST; at 10:1, 18 min; and, at 100:1, a mere 1.8 min. None of these is a great amount, and the researcher's ability to find structure in the sequence may be compromised in the following ways:

- The less ST analyzed, the more difficult it is to discern patterns and cause–effect relations at lower frequencies. If AT:ST is 100:1 and an analysis session 3 hr, then, in the 1.8 min of ST analyzed, any low-frequency regularities happening over many minutes will be missed. The analyst will have to detect these over several analysis sessions or deliberately move to a coarser grain of analysis to cover more ST in one session.
- Coding schemes can confine an investigator's attention to certain frequency bands just as much as to certain attributes of the recorded events. If a coding scheme focuses on low-frequency band event types, then regularities in higher frequency bands cannot be detected without returning to the raw data. If a coding scheme focuses on high-frequency band event types, then regularities in lower frequency bands can be detected (given adequate data) only if those regularities are of a kind that can be inferred or computed from events being coded in the higher frequency band.
- If the absolute quantity of ST that must be analyzed is large, such as 50 hr, then even relatively low AT:ST ratios will lead to prohibitively long ATs. The investigator is forced to abandon the full analysis or to radically alter the style of analysis.

So far, we have discussed rule-based linear coding that uses approximately the same AT:ST ratio throughout a session. The problems this approach poses help to explain why many experienced ESDA practitioners delay using a coding scheme, change their AT:ST ratio at critical points in their analysis, and avoid working linearly. In the initial stages, an analyst may continually move among event types, themes, and frequency bands—we call this *survey-based nonlinear coding*. In the later stages, the analyst may concentrate on chosen parts of the sequence using a style of analysis more like rule-based linear coding. The size of the combined AT:ST ratio from the survey-based and rule-based parts of the analysis may still pose practical problems if the analyst has to examine a specific amount of ST. However, an initial survey-based style of analysis opens the possibility of noticing regularities within different frequency bands—in a word, it is more *exploratory*. We now turn our attention to operations that support the exploration of observational data.

4. EXPLORATION AND THE EIGHT Cs

4.1. The Legacy of EDA

Many HCI researchers have commented on the relevance of Tukey's (1977) EDA for the type of observational data analysis they practice, but the connection is seldom taken much further (Mackay, 1989; Roschelle, Pea, & Trigg, 1990; Sanderson, James, & Seidler, 1989). We provide some background information about the ideas behind EDA and then conjecture how the EDA philosophy might be extended to sequential observational data.

EDA is a simple, visual, but still quantitative approach to data that allows an investigator to achieve a richer qualitative understanding by "looking at data to see what it seems to say" (Tukey, 1977, p. v). EDA helps researchers to generate hypotheses about what might be happening in a data set–rather than to test hypotheses, which is confirmatory data analysis. As a philosophy of data analysis, EDA is based on three principles (Hartwig & Dearing, 1979; Tukey, 1977):

1. *Continual openness and reexpression.* Instead of immediately imposing a model on the data, which might conceal important details, EDA analysts try to find patterns in the data and to describe them with simple summary statistics. It may take several iterations before the analyst achieves a satisfactory summary or "smoothing" of the data. The so-called smooth part of a data set is the variability that the analyst has accounted for so far, whereas the "rough" part is the variability that remains unexplained. Reexpressions or transformations of the data are essential for smoothing because they help the analyst see new patterns.
2. *Initial skepticism.* Because EDA analysts assume that there is no uniquely correct numerical summary of a data set, they are very skeptical of initial numerical summaries. Numerical summaries and smoothings are continually tested against the raw data to make sure they represent the data adequately. To help detect patterns and search for data points that do not fit the smooth part (outliers), EDA analysts rely heavily on visualization. Moreover, EDA analysts prefer statistics that are more sensitive to the bulk of data points ("resistant" statistics) rather than to outlying data points, unlike least-squares methods.
3. *Exploration versus confirmation.* EDA distinguishes exploratory from confirmatory data analysis but always maintains a productive interplay between them that is complementary rather than antagonistic. By supporting data exploration, EDA helps researchers generate hypotheses. These hypotheses can later be tested with formal confirmatory procedures using inferential statistics.

Although EDA has dealt with time-series analysis, in general it has been concerned more with static data sets than with dynamic sequential data sets. When noting the relevance of EDA for observational data analysis, HCI researchers sometimes overlook the fact that EDA has evolved into a highly sophisticated branch of statistics, complete with equivalents for a wide range of conventional statistics and its own modeling techniques (Leinhardt & Wasserman, 1979). We emphasize strongly that EDA's relevance for ESDA and HCI lies in philosophy rather than in technical details. Our goal is not to develop quantitative EDA equivalents for sequential statistics but instead to examine how the exploration of observational records can lead to strong, well-grounded statements in a realistic amount of time. What operations might support this?

It should be possible to adopt the general philosophy of EDA for sequential data, but many questions emerge. What is the smooth part in a temporal pattern, and what is the rough part? What is the best way to retain contact with raw data when they exist in time and often in another medium, such as video or audio? What are the most effective data transformations when data objects are symbolic rather than numerical? What kinds of reexpressions might help an analyst see patterns occurring over time?

4.2. ESDA Smoothing Operations: The Eight Cs

Relating the terminology of EDA to the generic ESDA process in Figure 1, Transformed Products can be viewed as the result of smoothings of sequential data. The particular smoothing operations chosen will depend on the currently relevant Formal Concepts and Research or Design Questions. The rough part should reflect only those aspects that are currently considered irrelevant.

We conjecture that there are eight generic transformations that could be viewed as primitive smoothing operations in ESDA: chunking, commenting, coding, connecting, comparing, constraining, converting, and computing–the "Eight Cs." They appear to be fundamental to ESDA practice, regardless of tradition. We base this conjecture on the many examples and methodological descriptions of ESDA that are now available; on the features researchers seek in general application software to help them analyze their data; on the features currently being built into ESDA software support; on preceding analyses (Mackay, 1989; Pitman, 1985; Poltrock & Nasr, 1989; Roschelle et al., 1990); and, last, on our own experiences as ESDA practitioners and software developers (Fisher, 1988, 1991; Sanderson, James, & Seidler, 1989; Sanderson et al., in press). In this section, we provide an abstract description of the Eight Cs; later in the article, we provide more concrete examples. For a complete example of the Eight Cs in action, see Sanderson's (1994) discussion of Rasmussen and Jensen (1973).

Figure 4 illustrates the Eight Cs. The icons next to each heading are shorthand forms that are used in later diagrams. Each set of small, stacked boxes represents a series of data elements at the highest frequency being considered, with time (t) running downward. Each "C" represents a class of operations that can take place on these data elements (and on transformations of them). As we see later, certain traditions–and certain techniques within each tradition–emphasize some Cs at the expense of others, but most of the Eight Cs are usually needed to describe how inference is supported in ESDA.

Chunks are aggregates of adjacent data elements that the analyst views as cohering. Chunking tends to be used most heavily at the outset, when analysts are faced with an unstructured mass of events whose only connection with one another is their temporal relation. For example, the major phases and phase transitions of a sequence may be drawn on a transcribed verbalization, or tags made on a videotape. Chunking is often hierarchical–data elements being grouped into chunks, chunks into larger episodes, and so on. Because chunking is effectively low-bandpass filtering, what emerges is a low-frequency band summary of the temporal structuring of a sequence.

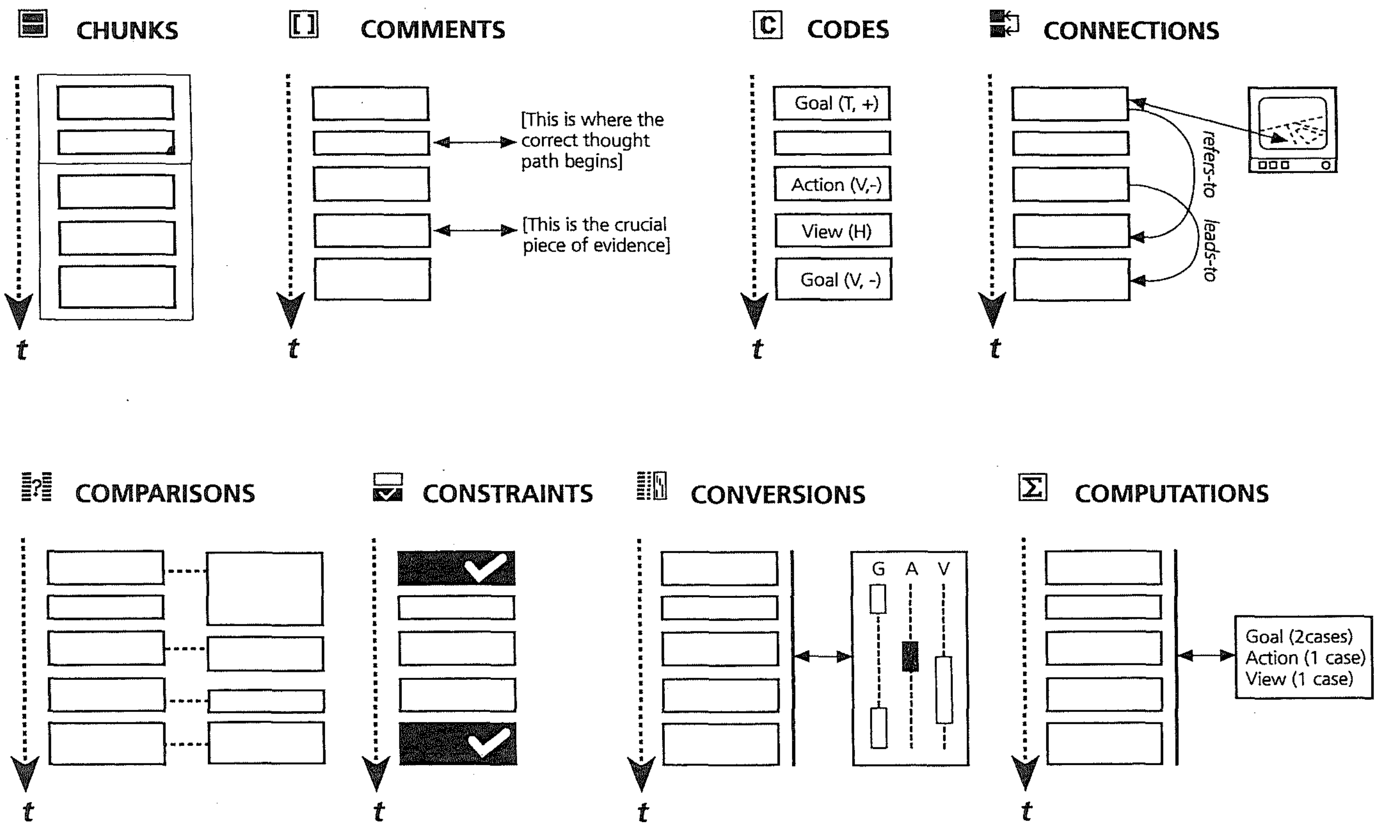
Comments are syntactically unstructured notes that are linked to any part of an analysis. Figure 4 shows comments linked with particular data elements, but they could just as well have been linked to products such as chunks and conversions. Comments may be summaries of what the analyst sees, or memos of theoretical reflections. Commenting often takes place alongside chunking.

Codes are syntactically structured labels that are usually linked to data elements or chunks. Figure 4 shows a highly structured predicate coding notation; however, codes are often simply keywords or combinations of keywords (see Sanderson, 1994). Codes spring from many sources, one being the comments made previously in analysis. As we see later, codes are used extensively in some kinds of ESDA but hardly at all in others.

Connections represent either the relations between ESDA products of a similar type or links between qualitatively different ESDA products that are nonetheless based on the same data elements. First, because of the situated, event-driven, and often haphazard nature of real-world events, topics, intentions, and themes do not necessarily emerge in a linear, temporally contiguous way but instead emerge in nonadjacent parts of a record. When using paper and pencil, analysts often draw arrows between different parts of their data in an attempt to draw out the more complex nonlinear structure. Second, and more mechanically, connections also represent links between qualitatively different products–such as between a data element and a video segment to which it is connected through time, or between a graphical element and a related comment.

Comparisons allow analysts to judge the effects of certain factors on sequential data. First, comparisons between the comments, codes, or

Figure 4. The Eight Cs: ESDA operations, or transformations, performed on Logs and Recordings to induce Statements.



chunks produced by several analysts may reveal differences in interpretation. Such comparisons are important in many ESDA techniques but for different reasons, as is seen later. Second, an analyst may wish to compare different parts of one subject's data or to compare different subjects' data, possibly collected under different conditions. Third, if a model or standard exists that predicts what should be seen, such as a cockpit check list or a computational model of HCI, then observed events must be compared with predicted events. Analysts can compare raw data, statistical summaries, or graphical reexpressions of the data.

Constraints represent a selection of part of the data for further transformation and a temporary exclusion of the remainder. In contrast to chunks, constraints are usually aggregates of nonadjacent data elements. Constraints allow the analyst to focus on certain products–events, chunks, comments, codes–without the distracting presence of other material. The analyst constrains by specifying Boolean combinations of the properties of products or by specifying temporal relations between sequentially ordered products. When combined with other operations, constraining is immensely powerful and is central to many important ESDA activities. First, constraints help ideas become better defined and properties of the data more consistently identified. An analyst may restrict a display of data to all video elements, data elements, or chunks that have comments linked with them that include certain keywords, or the analyst may compare all products encoded a certain way to check that they represent the same idea. Second, constraints are essential for certain types of querying and exploratory analysis. In a study of meetings, for example, an analyst may select lulls in conversation, focus on the participant who usually broke the silence, and determine the relative effectiveness of that participant's different communicative strategies.

Conversions are data transformations that allow new patterns to emerge. Conversions within a given ESDA product type can include converting coding done so far to a new scheme, possibly combining codes, or changing the resolution of a time-line graphic. Some properties are emphasized over others, but the ESDA product remains the same. Conversions across ESDA product types might include transforming alphanumeric data into a time-line display form or even translating it into auditory form. Conversions promote changes in the frequency band of analysis or a better thematic focus within a frequency band.

Computations refer to the broad range of formal procedures for analyzing sequential data that different ESDA techniques offer. As we see later, computations can refer to statistical operations typical of the behavioral tradition or to the coding and running of symbolic process models often seen in the cognitive tradition. Computations usually suggest confirmatory rather than exploratory data analysis, but they can equally well become powerful tools for exploration.

5. ESDA TRADITIONS AND TECHNIQUES

We have now established some important foundations for thinking about ESDA–some of which are reflected in Figure 1. We have defined ESDA and examined the types and frequency bands of events typically observed; we have discussed AT:ST ratios and identified important influences on them; and we have examined the relation between EDA and ESDA and conjectured that the Eight Cs are primitive ESDA smoothing operations. However, an important element is still missing from our discussion of ESDA–the role of Formal Concepts. As Figure 1 shows, Formal Concepts guide the researcher from Research or Design Questions to Statements. They shape how a research or design question is phrased, they point to the form statements should have to carry weight, and they indicate the types of analytic operations and Transformed Products that are appropriate along the way.

Most importantly, formal concepts arise from intellectual traditions. By a *tradition*, we mean a collection of research practices that have inherited at least some common fundamental assumptions about what are appropriate research questions, how such questions are operationalized and examined, and the forms an adequate answer can take. Traditions provide analysts with shared background assumptions about procedure and interpretation that are relatively seldom made explicit. We propose that there are three traditions of research most relevant for the use of ESDA in HCI–behavioral, cognitive, and social. If we are to understand how ESDA is used in HCI, to appreciate how ESDA techniques are similar and how different, and to speculate how ESDA techniques might borrow profitably from one another in the service of HCI research and design, the assumptions of these three traditions must be made explicit.

Figure 5 summarizes the characteristics of these three traditions. The first column lists the questions emanating from the Formal Concepts node of Figure 1, the second column translates the questions into more technical ESDA dimensions, and the remaining columns summarize how these dimensions are handled in each tradition. In each section that follows, we first discuss these characteristics. Then, for each tradition, we present one ESDA technique (or family of techniques) as an example, using expansions of Figure 1 to show how this technique works as a process. We then outline how each tradition and its techniques have been used in HCI and how they handle some of the ESDA issues already identified.

The descriptions that follow only approximate current practice and should not be viewed as definitive or prescriptive. Moreover, we emphasize that we do not view the behavioral, cognitive, and social traditions of ESDA as a historical progression from molecular to molar concerns, despite much talk about a recent "turn to the social" in HCI (R. J. Anderson, Heath, Luff, & Moran, 1993; G. M. Olson & J. S. Olson, 1991).

Figure 5. **How ESDA dimensions arising from generic ESDA questions are handled under the behavioral, cognitive, and social traditions of ESDA.**

ESDA Question	ESDA Dimension	Tradition: Behavioral	Tradition: Cognitive	Tradition: Social
"What's the issue at hand?"	Investigative approach	Use scientific method, often hypothetico-deductive tests, and achieve objective results	Model cognitive processes within the individual over time	Provide accounts of social phenomena; empirical, ethnographic, strong in inductive methods
"What should be observed?"	Setting	Field settings sought to ensure ecological validity; laboratory sometimes used	Systematic laboratory investigations, but also applied field settings	Field observations, emphasize value of observer's participation
	Sampling	Sampling theory and measurement theory used for subject and code selection	Ideally as for behavioral; detail needed often constrains sampling adequacy	Cover representative situations, pursue themes, artifacts or agents; seldom statistical sampling
	Focus of analysis	Objectively identified behavioral observables	Individuals' verbalization and action as evidence for cognitive processes and structures	Social interactions, communicative devices; utterance, gesture, action
"What operations should be done?"	Coding and description	Formal encoding using standardized terms; concern with reliability and objectivity	As for behavioral; debate whether use of context constitutes bias; induction often needed	Interpretation parallels encoding; participant involvement; use of multiple interpretations
	Means of analysis	Sequential, nonsequential statistical methods	Intensive assessment of goodness of model fit for a few subjects	Emphasis on qualitative rather than quantitative analysis; ethnographic
"What's an acceptable type of answer?"	Sources of rigor	Quantitative; concern with replicability and generality of results	Relation of account to extant models of cognition, computational adequacy	Qualitative; plausibility and robustness of account; finding one sound interpretation from many

Instead, HCI researchers using ESDA need to draw upon many different approaches if they are to pull events of different types and frequencies together in order to enrich and disambiguate one another.

The functionality of ESDA software tools often suggests how central certain ESDA practices are within particular traditions, so we provide references to software tools where relevant. For fuller surveys of ESDA software tools, see Harrison (1991), Ritter (1992), and, especially, Sanderson (1994), who classified ESDA tools according to their original purpose.

5.1. The Behavioral Tradition

Typical Characteristics

The behavioral[1] tradition is a distinct approach to the study of behavior that emphasizes objectivity and quantitative analysis. Observation is a well-defined approach within this tradition; observational practices in ethological, comparative, developmental, social, clinical, and human factors psychology go back many decades. A useful history of behavioral research in psychology, anthropology, and ethology can be found in Scherer and Ekman (1982), and thorough methodological treatments have been offered by Bakeman and Gottman (1986), Gottman and Roy (1990), Sackett (1978), Suen and Ary (1989), and van Hooff (1982). The behavioral tradition encompasses such research and design enterprises as human factors task analysis, method–time analysis, types of usability testing that focus on objective figures of merit or experimental contrasts, and theory development that relies on statistical analyses of encoded data, particularly using sequential data analysis techniques.

The typical characteristics of the behavioral tradition are shown in Figure 5. Overall, the goal is to achieve results that are as objective, replicable, and general as possible, given the conditions of observation. Thus, there is a formal commitment to the classical scientific method. Behavioral researchers with a strong motivation to study behavior as it unfolds naturally choose field settings for their observations, whereas others prefer to observe under laboratory conditions. Researchers rely strongly on sampling and measurement theory in order to handle issues such as time sampling, choice of subjects and material to analyze, development and standardization of codes, promotion of reliability, and management of observer bias. The focus of analysis is usually behavioral observables that can be identified objectively. Encodings are generally syntactically simple reexpressions and summarizations of raw sequences,

1. By *behavioral,* we do not imply a commitment to the behaviorist position but instead simply the adoption of the objective and quantitative approach outlined in this section.

filtered through a theory or viewpoint. Considerable emphasis is laid on finding (or developing) a coding scheme that will, when submitted to various statistical manipulations, speak reliably and validly to the research or design issues at hand.

In the behavioral tradition, analysis generally involves both descriptive and inferential statistics. Although nonsequential statistics are sufficient for many analytic purposes, a wide range of statistics preserving information about sequential relations has been developed. Thus, the sources of rigor for the behavioral tradition lie in collecting enough data to satisfy the assumptions of underlying statistical models and in finding appropriate ways to aggregate data so that statements with demonstrable generality can be made.

Example: Hypothesis Testing With Observational Data

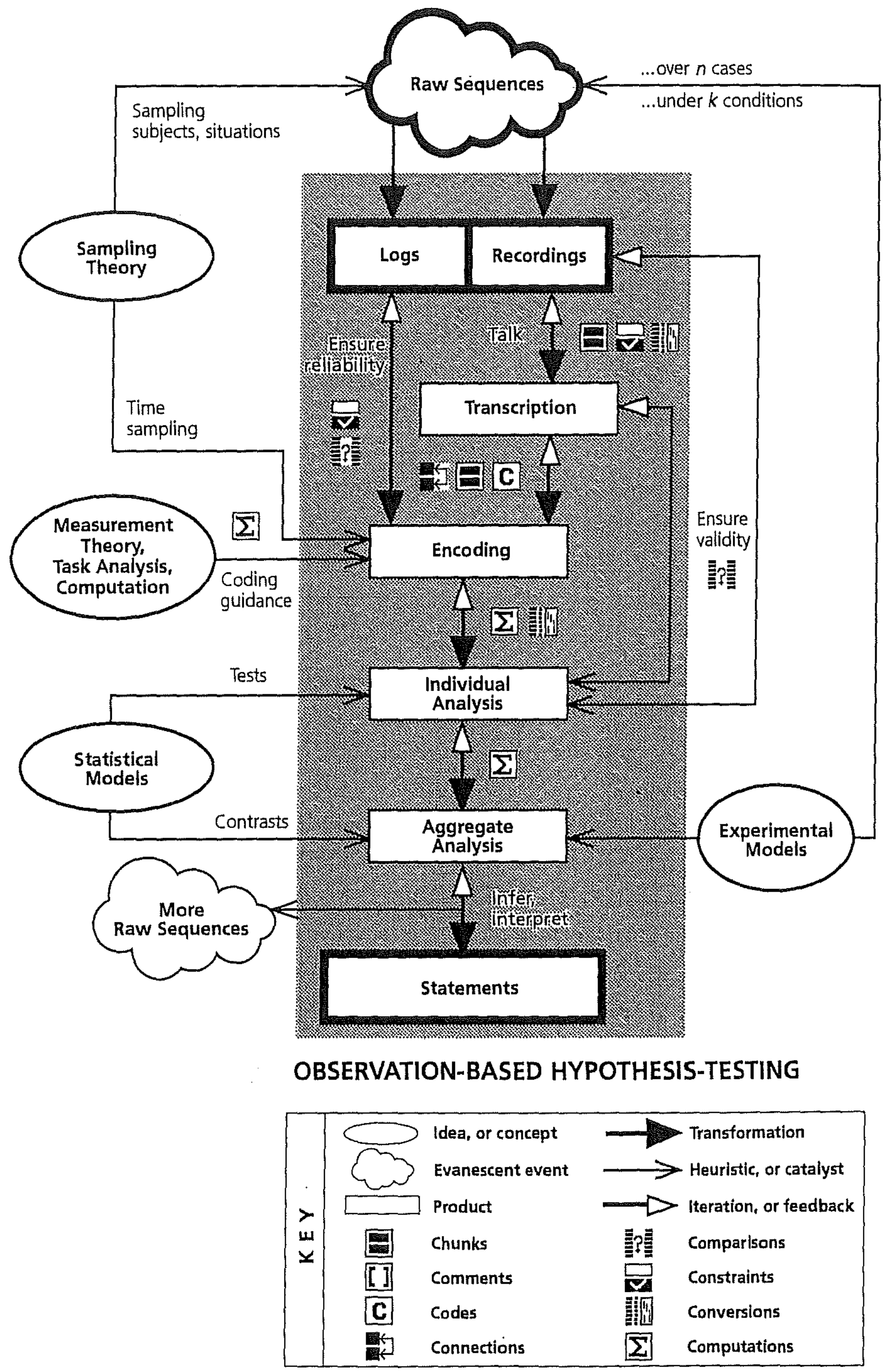
The use of statistical models in testing hypotheses based on observational data is an important behavioral technique, and our description here is based on many standard references (Bakeman & Gottman, 1986; Gottman & Roy, 1990; Suen & Ary, 1989). Suen and Ary (1989) provided a formal description of quantitative observational research:

> In a quantitative observational study, a complex phenomenon, behavior or otherwise, is first conceptually reduced to a number of *measurable* and observable behavioral variables. These variables are defined and a system for measuring these variables is determined before actual observation is done. A hypothesis that tentatively describes the relationships among these behavioral variables is then postulated. These relationships, if confirmed through empirical behavioral data, provide a quantitative description or explanation of the phenomenon. These defined behavioral variables are then observed to gather information or data of their quantities. Statistical analyses are performed on these quantitative data to determine whether evidence of the hypothesized relationships indeed exists. The hypotheses are either confirmed or disconfirmed. (pp. 5–6)

The commitment to the scientific method emerges in two places. First, sampling and measurement theory guide the analyst's decisions as data are being prepared for analysis, so that data conform to the assumptions of statistical models. Second, statistical and experimental models guide data analysis.

Preparation of Data for Statistical Analysis. Figure 6 shows the transformations, products, and sources of input involved in quantitative observational research. The Research and Design Questions node has been removed to simplify the diagram, and icons representing the Eight Cs have been placed on the appropriate arcs. *Sampling Theory* guides the choice of Raw Sequences for observation, pointing to appropriate subjects

Figure 6. **Hypothesis testing within the behavioral tradition of ESDA.**



and situations. Aspects of these sampled Raw Sequences that are likely to speak to the question at hand are captured in Logs and Recordings. There may be automatic capture of MKEs, eye-movement capture, audio and video recordings, or any combination of these. Sanderson (1994) has surveyed ESDA software tools that support data capture either by allowing the automatic capture and logging of events or on-line human coding of events, and many HCI-oriented tools with this capability were discussed in a recent special issue of *Behavior and Information Technology* (Nielsen, 1994).

When used in the behavioral tradition, verbalizations tend to be the type that individuals or groups would naturally make when performing a task, rather than being specially elicited from them, as in the cognitive tradition. Determining the content of verbalizations and making qualitative assessments about their meanings in context requires human judgment, which may be aided by having a *Transcription* available, creating a transitional step between Recordings and *Encoding*.

The goal of Encoding is to produce objective data that can be analyzed statistically; guidance comes from Sampling Theory and *Measurement Theory, Task Analysis, Computation*. To a limited extent, codes can be generated automatically. The logs produced by import, capture, or on-line human coding may serve adequately as encodings in their own right. However, some preprocessing may be needed to establish the appropriate grain of description for a meaningful statistical analysis. First, Sampling Theory can help determine how the temporal aspects of events should be represented. The options are to store event onsets only, to store event onsets and offsets, or to sample events at some chosen frequency at distinct points in time or across time intervals. Given basic event-onset and event-offset times, some software allows analysts to move among these possibilities as needed (Sanderson et al., in press). Measurement Theory provides standards to which codes should conform. For example, codes may have to make up a mutually exclusive and exhaustive set (Bakeman & Gottman, 1986) and be relatively few in number if data sparseness is to be avoided. Task Analysis is a detailed prior examination of the target domain by the researcher. It defines how the functional context of behavior gives rise to certain tasks (Kirwan & Ainsworth, 1992) and thus produces a language for describing these tasks as events and actions. Computation can be used to generate codes from data already obtained using techniques such as cluster analysis (van Hooff, 1982) or repertory grid analysis (Boose, 1984). Alternatively, computation can help a researcher find an appropriate semantics and frequency band for analysis through grammatical rewrite rules: Constraints are imposed and chunks created at the desired level through computation (G. M. Olson, Herbsleb, & Rueter, 1994).

In most cases, the human must do the encoding. Gesture must be encoded from video recordings, although automatic gesture-parsing is in its early stages (Bolt & Herranz, 1992). Verbalization can be encoded from recordings or transcriptions. Digitized audio signals can provide automatic information on vocalization initiation and termination times from different sources that, for some research questions, may include enough information for the results to be adequate for further analysis (Sellen, 1992), but inferring meaning is still a human task.

Finally, the connection from Encoding back to Logs and Recordings in Figure 6 represents the importance of achieving consistency in encoding within and between analysts. This comparison is often helped by calculating interobserver or intraobserver reliability. In the behavioral tradition, the most usual measure is Cohen's (1960) kappa, which tests whether agreement between encodings is greater than would be expected by chance (see Sanderson, 1994, for software that offers kappa). Alternatively, Cronbach's generalizability theory can be used to see whether encodings distinguish the variables that should be distinguished, even if maximum reliability is not attained (see Suen & Ary, 1989).

Statistical Analysis. Figure 6 shows how individual and aggregate analyses are performed. The *Individual Analysis* node shows that quantitative analyses (computations) can be conducted at the intrasequence level to obtain summaries of an individual sequence. A researcher may be interested in whether a person tends to refer to on-screen help or to consult a hard-copy manual. The number of events of each type can be counted, and a statistical test can be performed, to see if there is a significant bias toward on-screen help or the hard-copy manual. Alternatively, summaries of individual analyses can be combined into an *Aggregate Analysis,* in which analysis is conducted either at the intersequence level to examine replicability across sequences (do more people consult on-line help than would be expected by chance?) or at the intercondition level to examine the effect of independent variables (does reference to on-line help depend on expertise?). Figure 6 shows the results of individual analyses being fed forward to aggregate analyses, over n subjects. Here, experimental models seeking contrasts between k conditions may be used. Statements emerge from the way statistical results speak to the original questions.

Statistical analyses can be nonsequential or sequential. Content analysis–counting the frequency of events of interest or combinations of them–is the most common nonsequential statistic used. It has a long tradition in the social sciences and humanities as well as in the behavioral sciences, where it is supported by sophisticated software packages. It is almost universally supported by ESDA software tools (see Sanderson, 1994), and analyses of the durations of events are also widely supported.

However, the real power of the quantitative approach to ESDA lies in the use of sequential statistics, particularly for individual analyses. Sequential statistics find patterns in data sequences by seeking statistical dependencies between events over time, and they offer a window onto cause–effect relations. Sequential data analysis is a vast and fascinating topic, but only an extremely brief survey can be offered here. Gottman and Roy (1990), G. M. Olson et al. (1994), and van Hooff (1982) have offered fuller treatments.

The simplest approach to analyzing sequences is to look at transitions from one or more events to the next event and to determine whether the probability of this transition is greater than would be expected by chance. The full apparatus of Markov analysis can also be applied to explore dependency in behavior sequences by testing for the significance of longer and longer chains of events until no further variance is accounted for (Gottman & Roy, 1990; van Hooff, 1982; for ESDA software incorporating these techniques, see Sanderson, 1994). A similar approach, again based on transition matrices, can be taken using Shannon information theory to test for the information transmitted by longer and longer sets of prior events to a target event (Gottman & Roy, 1990). Other sequential data analysis techniques are similarity-based measures, such as the Pathfinder-based PRONET (Cooke, Neville, & Rowe, in press) and formal time-series analysis (Box & Jenkins, 1976).

However, data can become unacceptably sparse when higher order transitions or low-frequency band events are pursued. Lag sequential analysis (LSA) can help under such circumstances (Allison & Liker, 1982; Faraone & Dorfman, 1987; Sackett, 1978). Instead of seeking strict dependencies in sequences of adjacent events, LSA seeks dependencies between events that are separated by a constant number of other events. In EDA terms, with LSA the rough part in a sequence is less likely to get in the way of seeing temporal relations than it is with transitions, and the analyst can discern loose, rather than strict, temporal relations in a sequence. LSA is used widely in behavioral observational studies (Bakeman, 1983) and is increasingly represented in ESDA software used for HCI (Noldus, 1991; Sanderson et al., in press; Tapp, 1994). Other techniques useful for finding looser patterns in data are the analysis of cycles (Fisher, 1991), the search for maximal repeating patterns that indicate behavioral subroutines (Siochi & Hix, 1991), and the search for analyst-defined sequences of events or codes (G. M. Olson et al., 1994; Randle & Szostak, 1993).

The latter techniques provide the most promise for creating strong and expressive exploratory methods for temporal sequences. Regular expressions provide a possible foundation, but they are not sufficiently expressive to handle complex temporal relations among different streams of events. A fuller language is needed, perhaps sharing some of the features of the User Action Notation (Hartson, Siochi, & Hix, 1990). Mainzer's Pattern Expression Language (cited in Sanderson et al., in press) offers some of the required features, integrating temporal operators with regular expressions and database functions.

HCI and the Behavioral Tradition of ESDA

Examples in This Special Issue. Vortac, Edwards, and Manning's (1994) work on air traffic control (ATC) automation offers a good example of how the behavioral tradition can be used in HCI. Vortac et al. argue that modules of activity should not be fractionated when ATC operators transfer from current to new computational support for ATC, and they base their argument on a careful, objective, observational study typical of the behavioral tradition. They develop a set of coding categories that is large enough to represent the ATC task adequately but small enough to keep the chosen sequential data analysis technique (Pathfinder) tractable. Vortac et al. promote reliability by using two parallel encoders in a richly instrumented data-capture environment. They interpret the results of their Pathfinder analysis–a network of validated transitions between controllers' actions–in the context of an experimental design contrasting individual and group and contrasting low- and high-load conditions, even though they do not perform statistical contrasts. Interestingly, Cooke et al. (in press) recently extended the use of Pathfinder beyond Vortac et al.'s action–action transitions to state–action transitions–thus representing human activity as a response to system state rather than simply as a chain of actions–and they proposed statistics for comparing Pathfinder networks.

Also in this special issue, G. M. Olson et al. (1994) use sequential data analysis and grammatical techniques in a program of research on the impact of technology on meetings. Their use of Computation to tune coding categories and to find an appropriate level of description is especially typical of the behavioral tradition. They initially adopted codes representing conversational topics from a design rationale diagram but successively refined the topics until there were 22 stable categories that could be applied reliably. LSA and partial grammatical parses were used in an exploratory fashion to find sequential regularities in topics discussed. The sequences were then summarized, using rewrites, at a level of abstraction that provided a good description of meeting structure and differentiated meetings conducted under different conditions. Such computational techniques are powerful for finding the most appropriate semantics and frequency band for analysis.

Further Use of the Behavioral Tradition of ESDA in HCI. In principle, the behavioral tradition of ESDA can encompass a wide variety of HCI questions and types of data. It offers theory-neutral procedures for analysis–requiring simply that data be collected and prepared in a standardized way and that they do not violate the assumptions of any statistical tests to which they are subjected. Some researchers actively encourage observational hypothesis-testing as crucial to the development of HCI theory. McGrath (1992) has encouraged CSCW researchers to perform systematic parametric studies of groups interacting with technology so that factors affecting quality of group interaction can be understood in the most general way. Moreover, objective behavioral observation can be useful for certain kinds of HCI usability evaluations (Sweeney et al., 1993). Frequencies, probabilities, and durations of different types of observed activity can serve as nonsequential "figures of merit" for comparing alternative designs, different groups of users (e.g., experts and novices), or varying conditions of work (e.g., constraints on communication or on the types of tasks posed). Moreover, repetitive patterns of interaction can indicate interface problems (Siochi & Hix, 1991) or functionally related activities (Vortac et al., 1994).

Frequency Band of ESDA Event Types for the Behavioral Tradition. At first glance, it may appear that the behavioral approach works best with high-frequency band event types such as gestures, eye movements, and MKEs, but this is not necessarily so. In HCI, we find statistically oriented behavioral studies being conducted on low-frequency band event types such as computer conferences, meetings, and projects. For example, in a study of asynchronous computer conferences lasting from a few days to several weeks, Weedman (1991) performed a content analysis of message transcripts in the "turn" and "topic" range of the event spectrum (see Figure 3)–coding messages according to whether or not they were task related and checking intercoder reliability. Furthermore, in computer-mediated synchronous meetings, Lea and Spears (1991) manipulated the degree of group identity and the physical proximity of group members to examine the effect on interaction quality. They extracted messages from the stored message log, coded them for content (noting intercoder reliability), and carried out analyses of variance to see if content varied under different conditions. Moving to even lower frequency band event types, meetings can become data elements in the study of long-term projects. Electronic calendars and software-event notification systems can be adapted for the capture and review of such events.

Boundaries of the Behavioral Tradition

The behavioral tradition of ESDA provides HCI researchers with Statements taking the form "There are n instances of x," "x takes t time," or "Statistically significant sequences s_1, s_2, ..., s_n exist," and statements comparing conditions can be made on an ordinal, interval, or ratio scale. In many HCI environments, such precision is valued. If firm criteria exist to judge the desirability of x in these cases, then such statements serve design in a relatively direct way. They serve theory construction in a less direct way by providing evidence for or against theoretical predictions.

At the heart of the behavioral tradition is the need to establish a clear distinction between exploratory and confirmatory phases of analysis, especially if data are being tested against statistical models. During the exploratory phase, a possible relation might occur to a researcher, who finds a way to operationalize it, collects positive and negative instances, performs statistical tests on the results, and, on that basis, changes the coding scheme. However, during the confirmatory stage, the coding scheme and the tests chosen are critical parts of the formal operationalization of the hypothesis. If the hypothesis starts to fail as new sets of data are examined, neither the coding scheme nor the tests chosen should be changed–otherwise, degrees of freedom will be lost.

Many researchers in the behavioral tradition emphasize the importance of an exploratory phase, particularly in new areas of investigation (Bakeman & Gottman, 1986; Suen & Ary, 1989). Their comments echo Tukey's (1977) concerns. Nonetheless, within the behavioral tradition of ESDA, exploration is often treated as an informal supporting practice, and little attention is paid to how to make it effective. Bakeman and Gottman (1986) warned:

> However, [exploratory hypothesis generation] needs to be carefully done by incorporating it in programmatic research that builds in replication. ... There is a danger in this hypothesis-generating approach, and this has to do with the temptation of not thinking very much about what one might expect, and instead looking at everything. ... Usually investigators who do not generate hypotheses at all will be overwhelmed by their data. (p. 17)

A balance is needed between endlessly exploring possible leads and prematurely moving into confirmatory activities that may degenerate into prolonged encoding efforts and fatuously precise–but ultimately uninformative–statistical tests. It is against the latter case that the most pointed criticisms of the behavioral tradition are leveled. When we present the social tradition, we examine how exploration might be structured more effectively.

5.2. The Cognitive Tradition

Typical Characteristics

Investigators in the cognitive tradition view human thought as being describable by a series of processes operating on information that is represented mentally in various ways (J. R. Anderson, 1993; Gardner, 1985; Newell, 1990; Posner, 1989). Even though the foundations of the cognitive tradition lie more in formal logic, cybernetics, and formal linguistics than in statistical concepts, precision and objectivity are valued as much as in the behavioral tradition, and cognitive theory can be developed through behavioral hypothesis-testing. However, if we examine research practices, we see that much cognitive research using ESDA has characteristics that justify treating it as a separate tradition. These research practices include cognitive task analysis, protocol analysis, discourse analysis, development and testing of symbolic computational models based on human data, and search for hierarchical structure using grammatically based approaches. Moreover, a distinct class of ESDA software support has emerged for building and testing cognitive theories based on complex sequential data (Ritter, 1992; Sanderson, 1994).

The principal characteristics of the cognitive tradition are shown in Figure 5. They emphasize the symbolic approach as practiced in the United States–rather than the neurally inspired approach (Rumelhart, McClelland, & PDP Research Group, 1986), whose practitioners typically have not used ESDA. Verbal and nonverbal behaviors are assumed to be products resulting from cognitive processes, and sometimes researchers try to model the operation of these processes within the individual over time. An influential (but certainly not universal) research practice is to operationalize these models as computer simulations that provide an output "trace" of the hypothesized products of cognitive processes that can be compared with a human trace. In other cases, researchers try to infer knowledge structures from the content of utterances or the presence or absence of actions under certain conditions–rather than from the order of outputs.

ESDA in the cognitive tradition is more typically performed in a laboratory setting than in the field, although many field studies have been conducted. Sampling generally proceeds as for the behavioral tradition, but there is a tendency to perform extensive analyses of the data from a few subjects or from only one subject rather than to search for generality across multiple subjects. Analysis and modeling are often so complex that only a few subjects' data can be analyzed.

Cognition has traditionally been treated as a phenomenon happening at–and therefore capable of being sufficiently modeled at–the level of the individual. Therefore, the focus of analysis is usually the activities and verbalizations of individuals. An important analytic assumption is that an individual's verbalizations validly reflect the information, or cognitive products, heeded in short-term memory and so can serve as evidence for underlying cognitive processes (Ericsson & Simon, 1993).

Coding and description conform to the same standards as for the behavioral tradition. However, these standards are more difficult to achieve because cognitive processes are unobservable and must be inferred from cognitive products. As a result, coding and analysis become intensely inductive and iterative (Ericsson & Simon, 1993; Newell & Simon, 1972; Ritter, 1992; Waterman & Newell, 1971). The analyst moves in an exploratory fashion between developing and changing encodings–testing the fit of data to a model, altering the model, and judging when a parsimonious scientific explanation has been achieved. Last, the source of rigor is how tightly an account explains a phenomenon in the context of extant models of cognition.

Example: Cognitive Process Modeling

Our description of cognitive-process modeling is based on formal treatments in Ericsson and Simon (1993), Newell and Simon (1972), and Simon and Kaplan (1989) and in the details of many studies performed in specific domains. It includes the development of noncomputational cognitive process models and the development, fitting, and testing of computational models. This process is shown in Figure 7.

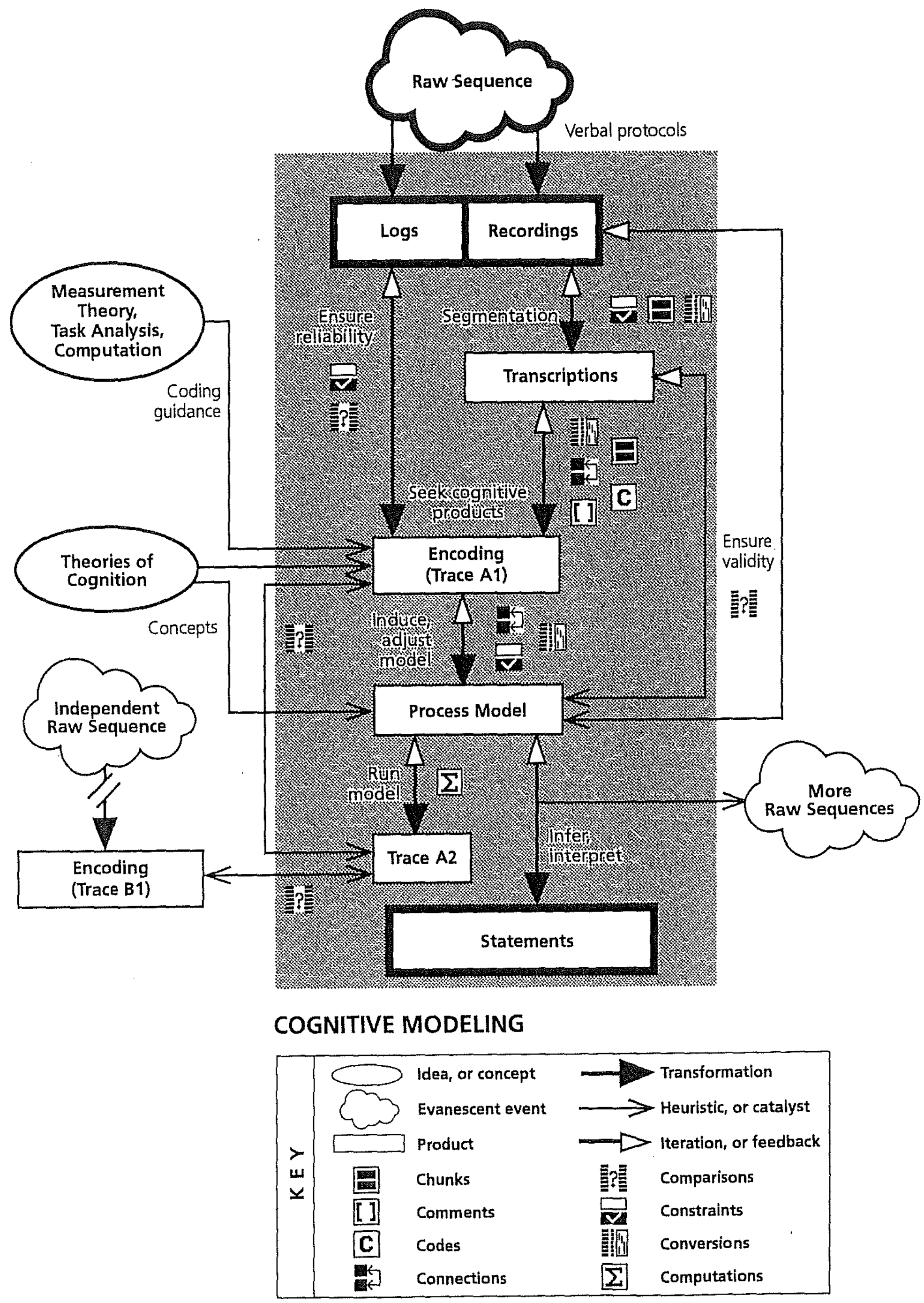
Because the cognitive and behavioral traditions share a concern for objectivity and precision, cognitive process modeling echoes many features of hypothesis testing with observational data, particularly in the preparation of Logs and Recordings for analysis. However, cognitive-process modeling places a greater emphasis on verbalization because of its role in helping the analyst infer unobservable cognitive process.[2] Subjects are often asked to think aloud while performing their task–producing a verbal protocol that includes far more than the verbalization required by the task alone. Because it is not a natural part of a subject's task, there is always concern whether verbalization will affect the way a task is performed (i.e., reactivity) and whether it accurately reflects cognitive products in short-term memory (i.e., validity) (Ericsson & Simon, 1993; Russo, Johnson, & Stephens, 1989).

Verbalizations are converted into Transcriptions early in the ESDA process. Transcriptions tend to focus on utterance content rather than on the specifics of delivery. They are usually divided into segments that each supposedly reflect an elemental cognitive process–this is a high-bandpass chunking process. Segmentation can be done manually, but some software tools support a first pass on the basis of prosodic or semantic markers (Bailey & Kay, 1987; Poltrock & Nasr, 1989). In contrast with what is done in the behavioral tradition, time sampling is seldom used. For many types of cognitive modeling, sequential order is more important than real time, and many software tools supporting cognitive modeling do not directly store event time-stamps (Bhaskar & Simon, 1977; Fisher, 1991; Poltrock & Nasr, 1989; Sanderson, James, & Seidler, 1989; Waterman & Newell, 1972).

Figure 7 shows that the formal concepts contributing to Encoding include *Theories of Cognition.* The goal is to describe cognitive products in a way that will most effectively point to cognitive processes. Therefore, encoding syntaxes are often more complex than those used in the behav-

2. The importance of verbal protocol analysis for this tradition is attested to by the fact that many ESDA software tools originating in this tradition have the letters *P* and *A* somewhere in their acronyms, such as PAW, PAP, SAPA, SHAPA.

Figure 7. **Cognitive modeling within the cognitive tradition of ESDA.**



ioral tradition. Ericsson and Simon's (1993) adoption of a relational syntax has been very influential. As a result, coding syntaxes are often multivalued: The analyst chooses a main relation (e.g., "goal") and fills in supporting arguments (e.g., speaker="Bill", type="prepare list") (Conti, Cuomo & Rogers, 1994; Fisher, 1991; Poltrock & Nasr, 1989; Sanderson, James, &

Seidler, 1989; Sanderson et al., in press). The concern for encoding reliability is shared with the behavioral tradition, and similar methods are usually used. However, Newell and Simon (1972) commented that, "in [their] judgment, the long-run solution to these questions of validity and reliability lies in mechanizing the procedure–in getting the human analyst out of the business of making the attributions and creating the web of inference" (p. 184). The early automatic protocol-analysis programs PAS–I and PAS–II (Waterman & Newell, 1971, 1972) were attempts to promote reliability through automation.

If no prior model or coding language exists, the analyst first uses survey-based nonlinear coding. Chunks, connections, and comments are often used. Preliminary codes are applied in an exploratory fashion, and analyses (computations and comparisons) may even be tried, after which the coding scheme is refined. Because the material of interest may not be "observable," AT:ST ratios can become very high.

After an encoding language has been established, it must be applied consistently to the data to produce a trace, so the analyst switches to rule-based linear coding. If the task produces many raw data points, the result (*TraceA1*) will be dense. As the number of data points increases, richer models can be developed and stronger tests performed of the adequacy of these models (Ritter, 1992). If multiple event types such as verbalization, eye movements, and actions must be considered together, then the conceptual density and the AT:ST ratio climb steeply.

Analysts in the cognitive tradition often try to keep dependencies between coded data segments as low as possible so as not to reduce the degrees of freedom in the data (Ericsson & Simon, 1993). Because dependencies often arise when context is used in coding, some ESDA software tools are able to present the analyst with randomly chosen segments for coding (Bailey & Kay, 1987; Ericsson & Simon, 1993; Sanderson, James, & Seidler, 1989). The off-line coding effort is intense in the cognitive tradition, so supporting tools have emphasized the ability to change encoding vocabularies rapidly and the use of power keys or macros for routine activities (Ritter & Larkin, 1994). Tools also support inference, which may involve making connections between disparate parts of the data, converting the data into different formats in order to allow patterns to emerge, and constraining transformations to certain types of statement or activity in order to pursue certain themes more clearly (MacWhinney, 1991; Ritter & Larkin, 1994; Sanderson et al., in press; J. B. Smith, D. K. Smith, & Kuptsas, 1993).

The major difference between the behavioral and cognitive traditions lies in the analysis stages. Figure 7 shows that the next ESDA product is a *Process Model*. The analyst must induce the relation of each elemental product in a sequence (each act, state, or verbalization) to some unknown whole that allegedly underlies it (a coherent cognitive process). Because there is no equivalent of the "automated" procedural rationality of statistics that reveals the whole picture, Statements succeed or fail more at the individual subject level, as represented in the sequence, than at the sample or experiment level.

The lower part of Figure 7 shows how cognitive process models are used. In the simplest but most common case, the Process Model is not implemented as a computational procedure, but inferences are drawn from it that lead to final Statements. Computational implementation is bypassed if the model results from a cognitive task analysis aimed at supporting design decisions, if the model is couched at too general a level of description for implementation, or if the model is to be tested in noncomputational ways, such as through tests of specific performance predictions. Some ESDA software tools automatically build graphical representations (conversions) of models (Pitman, 1985; Poltrock & Nasr, 1989; Ritter & Larkin, 1994).

However, if the process model has been implemented computationally (computation), it can be tested by comparing its output, *TraceA2,* with the encoded trace of the data on which it was based, TraceA1. The fit can be judged, and either the encoding or the model can be adjusted to improve the fit, as seems appropriate. The model's wider generality can be tested by seeing how well TraceA2 matches an independent data trace from another subject, Encoding (*TraceB1*), which has been coded with the same categories (see Figure 7). A thorough review of these methods is provided by Ritter (1992), and further examples with software support are provided by J. R. Anderson (1993), Bhaskar and Simon (1977), John (1994), Kowalski and VanLehn (1988), and Ritter and Larkin (1994). As Sanderson (1994) noted, ESDA software tools for cognitive process modeling share many features with knowledge elicitation tools used in expert system development.

As noted, large sample sizes are seldom used for cognitive process modeling. The problem is even more extreme when it comes to testing the predictions of computational models; Ritter and Larkin (1994) show that more than 50% of the studies they review used only one subject. An important reason is that the task of the analyst is conceptually very dense–more so than is usual for the behavioral tradition–and AT:ST ratios can become extremely high. However, if sequences from more than one subject are analyzed, and figures of merit are derived, then the confirmatory stage of hypothesis testing will apply, as for the behavioral tradition.

HCI and the Cognitive Tradition of ESDA

Examples in This Special Issue. The cognitive-process models most often associated with the uses of ESDA in HCI are based in production systems or grammars. Both usually reflect hierarchical cognitive control architectures (Card, Moran, & Newell, 1983; Green, 1991; Kieras & Polson, 1985; J. S. Olson & G. M. Olson, 1990) in which it is assumed that behavior is organized around achieving goals according to some plan. Moreover, both allow strong predictions to be made about expected sequences of actions or thoughts that can be tested against actual sequences. In this special issue, G. M. Olson et al. (1994) and Ritter and Larkin (1994) take different approaches to the challenge of modeling processes, emphasizing the role of exploration.

Ritter and Larkin (1994) show how their Soar/Model-Testing (SMT) software helped to evaluate a model of the cognitive processes of users browsing an unfamiliar database of programming-language commands (the Browser–Soar model of Peck & John, 1992). Specifically, SMT helped Ritter and Larkin compare subjects' behavioral traces with Browser–Soar's output trace. Apart from making it easier to align actual and predicted traces and providing algorithmic support for maximizing trace matches, SMT provided strong visualization tools for detecting trace mismatches and maintaining connections between different forms of data. Ritter and Larkin show that tools like SMT not only speed the evaluation of process models–lowering AT:ST ratios and making it possible to analyze more data with greater rigor–but they also improve the quality of an analyst's thinking. Promising themes emerge more quickly from the data and can be explored more readily. Although SMT was developed in the context of the Soar cognitive architecture and model-development environment (Newell, 1990), SMT's trace-comparison approach can be used to test any process model that can generate an output trace of human activity. John (1994) developed a similar approach to test GOMS (Goals, Operators, Methods, Selections) models, and Johnston (in Sanderson et al., in press) extended the approach to help compare different subjects' traces, compare different analysts' interpretations, and compare prescribed operating procedures with actual human performance.

G. M. Olson et al.'s (1994) ESDA studies of meetings were introduced under the behavioral tradition, but their use of grammars contributes to the cognitive tradition of ESDA even though the article's subject matter is more social than cognitive. They provide a particularly lucid exposition of how definite clause grammars (DCGs) can be used in an exploratory fashion to model any sequence. They emphasize that they do not aim for a complete parse of an entire sequence but instead use DCG rewritings in an exploratory fashion to summarize and characterize as much of a sequence as they meaningfully can. Their techniques can readily be adapted to represent cognitive processes within an individual.

Other Grammar-Based Work. Two further ESDA-based research programs that have used grammars to understand cognitive processes are those of Hollnagel (1979) and J. B. Smith et al. (1993). Working in the context of human interaction with complex industrial systems, Hollnagel used a rule-based, context-free grammatical framework to build a hierarchical competence model of human performance and contrasted three representations–parse trees, rewrite rules, and a formatted trace output similar to those seen in GOMS and Soar analyses. More recently, J. B. Smith et al. explored the effectiveness of different types of grammars for building cognitive models of how writers develop written documents. Their most recent models are based on ATN (Augmented Transition Network) grammars that are context sensitive and that allow flexible parse-tree structures. G. M. Olson et al. and J. B. Smith et al. share an appreciation for the open-ended exploratory aspects of using grammars as well as the goal of making the analysis of vast amounts of sequential data tractable by providing strong visualizations of transformed products.

Further Use of the Cognitive Tradition of ESDA in HCI. Models such as GOMS (Card et al., 1983) continue to be used to describe and predict the sequential aspects of human interaction with computers (Bovair, Kieras, & Polson, 1990; Irving, Polson, & Irving, 1994) and recently have been extended to deal with sequential dependencies between events or activities using critical path analysis (Gray, John, & Atwood, 1993). Sometimes predicted and obtained sequences are compared in great detail, whereas sometimes only aggregate measures are compared, such as time to completion (Gray et al., 1993). Now that there is a greater focus on HCI as dialogue (Grudin, 1990), such models are being extended to activities such as information retrieval and browsing (Hoppe & Schiele, 1992; Peck & John, 1992) in which users show less sequential structure in their behavior. However, hierarchical goal-driven models start to fail for certain types of design problem-solving or open-ended browsing in which goals may be ill defined and behavior becomes opportunistically rather than hierarchically organized in time (Guindon, 1990; Hoc, 1986).

The cognitive tradition of ESDA in HCI extends well beyond cognitive process modeling. First, ESDA has been used to identify and model knowledge that can be inferred from the content of utterances or from the presence or absence of certain actions under certain conditions (Gentner & Stevens, 1983; Kieras & Bovair, 1984; Sanderson, Verhage, & Fuld, 1989; Shrager & Klahr, 1986). In these cases, the fact that an Encoding is ordered in time is usually unimportant. The Process Model in Figure 7 becomes a *Mental Model, Device Model,* or *System Knowledge* and is tested by seeing how well it predicts qualitative aspects of human performance under transfer conditions–such as problems encountered when working with a new interface–rather than by comparing output traces. ESDA studies focusing on the structure of knowledge have led to the development of intelligent tutoring systems and interactive advisors and, in some cases, have evolved into fully confirmatory techniques (J. R. Anderson, 1993).

Second, ESDA techniques derived from the cognitive tradition are often used without a commitment to cognitive process modeling. Think-aloud protocols are used to elicit preferences, priorities, emotions, associations, and analogies–material that is distinct from the usual cognitive products but that has been useful in HCI testing and evaluation (Gould, 1988; Karat, 1988; Sweeney et al., 1993).

Third, cognitive models have been built using ESDA techniques but without the help of think-aloud protocols. In well-defined domains, nonverbal activity may be sufficient to constrain the form of a cognitive model (J. R. Anderson, 1993). Moreover, in studies of distributed cognition, normal task-related talk may be as effective for exposing knowledge and skills within a group as it is for coordinating the group's work (Hutchins, 1991; Woods, 1993).

Frequency Band of ESDA Event Types for the Cognitive Tradition. In information-processing models of cognition, the duration of significant elemental processes is claimed to range from .1 sec to just less than 10 sec (Newell & Card, 1985; see Figure 3). However, at slightly lower frequency bands, there are further cognitive processes that are important for HCI–particularly topic event types enduring from fractions of minutes to multiples of minutes. Much ESDA in the cognitive tradition appears here–especially verbal and nonverbal protocol analysis, discourse analysis, process tracing, and so forth, which focus on the frequency band within which knowledge is applied, understanding develops, and learning takes place. In even lower frequency bands, we encounter operations and projects, for which models of distributed cognition in which molar social and temporal factors are more important than molecular information-processing factors (Cicourel, 1990; Hutchins, 1995). In such complex low-frequency bands, hypothesis testing becomes intractable, yielding to observational studies and sophisticated modeling techniques (Green, 1991; G. M. Olson & J. S. Olson, 1991).

Boundaries of the Cognitive Tradition

In the cognitive tradition, the theoretical significance of a model comes from the degree to which it articulates with and extends established theories of cognition and its parsimony in relation to its breadth of prediction. Just as for the behavioral tradition, model development (an exploratory phase) should be kept separate from model testing (a confirmatory phase).

Although theoretical significance is often less important than practical usefulness in applied work such as HCI design, theoretical significance need not disappear when a researcher works on practical applications or under complex field conditions. For example, Woods (1993) proposed a cognitively motivated ESDA method for studying "cognition in the wild" (Hutchins, 1995). Woods encouraged researchers to design observational settings in which subjects will be "externalizing internal process or producing external signs that support inferences about internal [mental] workings" (Hutchins, 1995, p. 233). Researchers are encouraged to think simultaneously in the specific language of a domain and in the abstract language of cognitive theory and to design controlled contrasts into observational sessions wherever possible. Woods's method is best understood through examples of its use in HCI and human–machine systems contexts, such as human interaction with an intelligent advisor (Roth, Bennett, & Woods, 1987), physician interaction with intravenous control devices (Moll van Charante, Cook, Woods, Yue, & Howie, 1993), and pilot interaction with cockpit automation (Woods & Sarter, 1993).

5.3. The Social Tradition

Typical Characteristics

The social sciences–particularly sociology, anthropology, and organizational theory–are becoming increasingly important for the study of HCI (G. M. Olson et al., 1992). Here we focus on work influenced by the general orientation of naturalism (Blumer, 1954; Schatzman & Strauss, 1973) and, more recently, by interpretive sociology and anthropology (Schwandt, 1994). These orientations often use ESDA methods. For example, ethnographically motivated field observation and participant observation often result in sequential records–and recordings–that must be painstakingly analyzed (Kendon, 1982). It is also within these orientations that much influential current work in the social aspects of HCI is being performed, driven by the need to observe habits of work and the use of systems. Further, qualitative techniques within the social tradition may even help to overcome some of the more intractable problems of ESDA that cut across all traditions.

The third column of Figure 5 shows how the social tradition of ESDA differs from the other traditions. We emphasize that not all the characteristics we mention necessarily underlie every type of naturalistic social-research practice (for comparisons between different techniques, see Adler & Adler, 1994; Guba & Lincoln, 1994). Overall, however, the approach to research and design questions is to understand the social, organizational, and material worlds within which individuals and groups operate and to generate accounts of how individuals and groups use available resources to reach goals and make sense to one another. The observer's participation in the target setting–almost always a field setting–is often highly valued because the goal is to better understand situations from the domain practitioner's point of view, rather than to impose abstractions from an outsider's point of view. This approach is particularly associated with ethnography–the principal research technique of anthropology.

Researchers try to cover representative situations rather than conforming strictly to the requirements of sampling theory. Thus, material tends to be sampled on an ongoing basis–themes, artifacts, and agents are followed as they become relevant (e.g., Strauss, 1987). The focus of analysis is social interaction, which points to subject matter such as communication and group work but also to a wide range of analytic orientations. At one extreme, researchers place individual cognition in social context (Hutchins, 1995; Star, 1989); at the other extreme, researchers believe that inferring the mental processes of individuals is profoundly inadequate for understanding how meanings are conveyed because meaning is fundamentally social (Suchman, 1987; Winograd & Flores, 1986). Researchers taking the latter view reject the cognitivist notion that prior constructs such as goals and plans account strongly for behavior, and they prefer models of highly context-dependent or "situated" cognition (Clancey, 1993; Suchman, 1987; Whiteside & Wixon, 1987; for an overview, see Holstein & Gubrium, 1994).

Coding, when performed, is usually approached quite differently from other traditions. Codes tend to already reflect interpretation rather than be abstractions relying on processing for their meaning to emerge.[3] Moreover, social researchers recognize that multiple valid interpretations of observational data exist. Because researchers try to infer the meaning of raw events in the face of these possible interpretations, codes sometimes appear more like final research statements than data points (see, in particular, Strauss, 1987). Although the goal may be to uncover the internal organization of phenomena, interpretational bias is acknowledged and explored, rather than eliminated, because it is held that no neutral point of view can exist. Thus, there is room for greater involvement of the participants themselves in the interpretation and analysis of the data. Altogether, bias is often managed quite differently from how it is managed in the other traditions.

Finally, analysis is usually qualitative rather than quantitative, relying on search, collection, comparison, and interpretation rather than modeling and statistical testing. Many different approaches to qualitative data analysis exist (Denzin & Lincoln, 1994; Garfinkel, 1967; Geertz, 1973; Miles & Huberman, 1994; Strauss, 1987), reflecting different viewpoints on appropriate subject matter, methodology, and the role of hypothesis testing. Recent surveys of software supporting ESDA in the social tradition

---

3. Codes always reflect interpretation. The point here is that, in the sociological tradition, codes are closer to final statements than to encoded data. When codes "encode" raw data, statistical analyses or other further operations have to be performed before the meaning of the data is revealed.

can be found in Miles and Huberman (1994), Richards and Richards (1994), and Sanderson (1994).

Example: Interaction Analysis

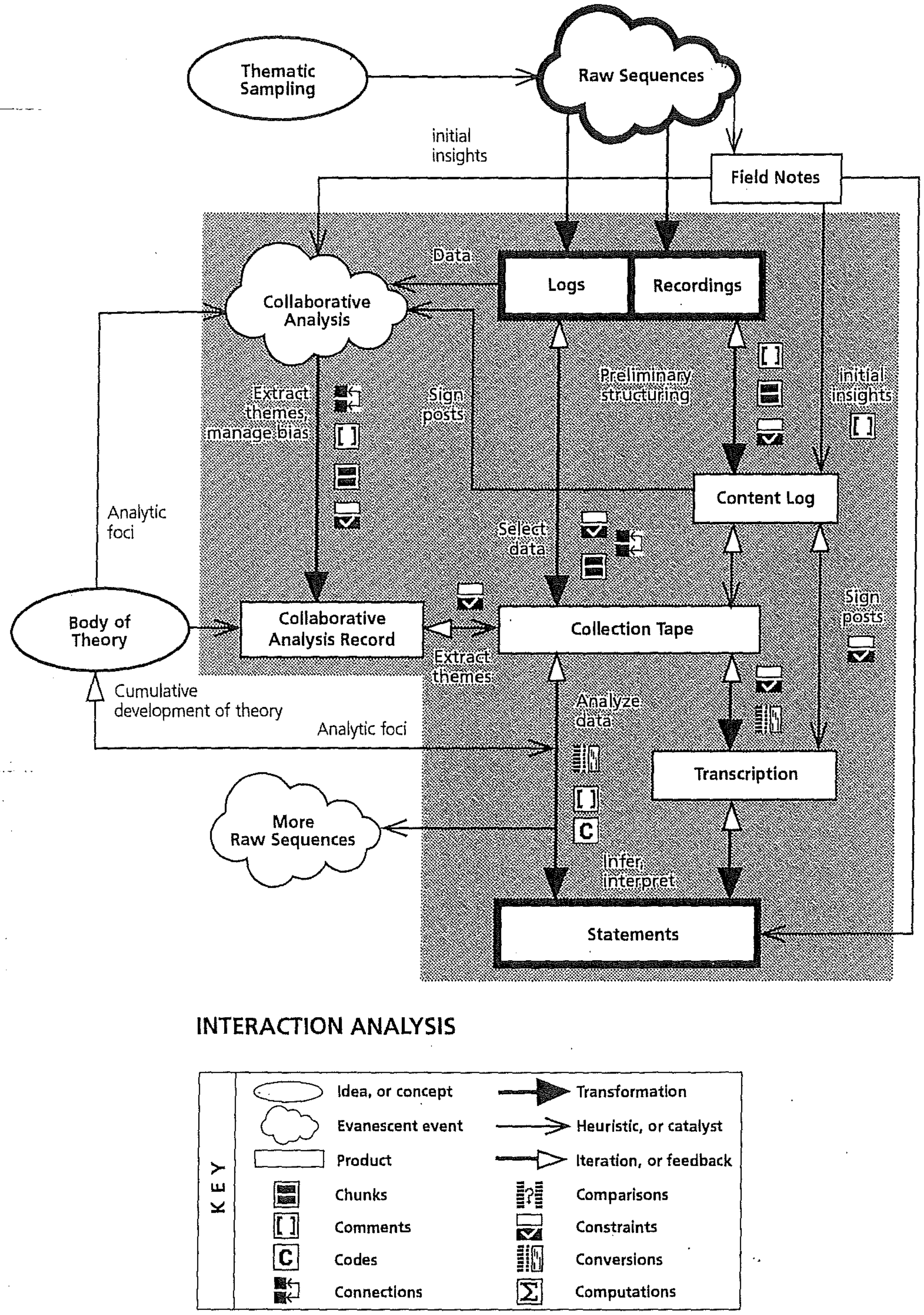
Interaction analysis is a collection of research practices for studying interactions among people. It has been used to study psychotherapy, learning, and health-care delivery but has recently been used within the HCI community to study interactions among people in work environments and among people and engineered systems. Interaction analysis focuses on how normal, unexceptional work practices are developed and maintained, rather than on exceptional episodes. The goal is to understand how people achieve mutual intelligibility in any natural or engineered environment. Such understanding requires inference, but strong emphasis is placed on grounding inferences in raw evidence–particularly evidence gathered through participant observation.

Although the term *interaction analysis* is often used generically, it also points to the practices of specific research communities. One important type of interaction analysis has emerged, centered on Xerox PARC, for studying the workplace (Jordan & Henderson, 1995; Roschelle et al., 1990; Suchman & Trigg, 1991; Tatar, 1989). Within this research community, interaction analysis is highly interdisciplinary. Although it is strongly influenced by ethnomethodology and associated with "situated action" theories, this version of interaction analysis is broader than conversation analysis. For example, Jordan and Henderson (1995) and Suchman and Trigg (1991) emphasized that it is often accompanied by ethnographic fieldwork, in which analysts themselves participate in the target work setting and use traditional ethnographic practices (e.g., note taking, examining artifacts). The need for such methodological eclecticism was also noted by Heath and Luff (1992b), who have studied gesture in workplace interaction.

The version of interaction analysis illustrated in Figure 8 is based on descriptions in Jordan and Henderson (1995), Suchman and Trigg (1991), and Tatar (1989). In practice, the sequence of events may not be exactly as in Figure 8, but the figure includes several possibilities that capture some of the characteristics of interaction analysis as these practitioners describe it.

In contrast to behavioral hypothesis-testing approaches, in interaction analysis a research question may be posed as much to generate hypotheses as to test them. The selection of Raw Sequences is driven by *Thematic Sampling* in which researchers focus on certain themes or dimensions of an interaction rather than use statistical sampling theory. For example, Suchman and Trigg (1991) noted that observation can focus on aspects of the work setting as a whole, on a certain person who has a particular role and viewpoint, on objects such as important documents, or on specific operations or

Figure 8. **Interaction analysis within the social tradition of ESDA.**



tasks. Here, interaction analysis inherits ethnographic practices. It also inherits from ethnomethodology the idea that meaning and proof are local, residing only in specific instances and events. As a result, a single data sequence can provide an existence proof for some cultural or social process at work, making statistical replication unnecessary and undesirable (Benson & Hughes, 1991).

Raw Sequences are not only captured in Logs and Recordings (typically video) but also in *Field Notes,* again reflecting ethnographic practices. The analyst then typically makes a *Content Log,* which is a rough temporal map of the video record and summary of its structure and highlights (for software support, see Mackay, 1989; Roschelle, 1994; Roschelle & Goldman, 1991). Creating the log involves extracting what is pertinent (constraints), discerning low-frequency structure (chunks), and making comments. The log can be used to raise issues, to indicate events of interest, to point to places where more data might be needed, and to suggest themes, categories, and strategies for analysis (Suchman & Trigg, 1991).

The recordings and content log are often brought into a *Collaborative Analysis* session, in which a group of researchers identifies promising themes in the data, corrects factual misperceptions, and manages bias by exploring multiple possible interpretations. Analysis is done collaboratively, often by a group of researchers, but sometimes also including domain practitioners, so that the "native" point of view is represented. Analysts guide themselves into the material using "analytic foci," which are interactional phenomena that have been found to be effective catalysts for deeper understandings (Jordan & Henderson, 1995). Analytic foci are suggested by the relevant *Body of Theory,* and they provoke analysts to ask questions such as, "How are activities initiated?" "How is the temporal patterning of work produced?" "How do participants detect and correct misunderstandings?" "How do participants use artifacts to support their work?" The analyst retains an audio log of the collaborative analysis session–the *Collaborative Analysis Record*–and may edit significant parts of the video data into a separate working *Collection Tape.* Collection tapes store the critical data from which generalizations are drawn, and they can be assembled and played in various orders. Constraining, chunking, and connecting are crucial to these activities. To differing degrees, ESDA software tools support the preparation of collection tapes (Bevan & Macleod, 1994; Goldman-Segall, 1992; Mackay, 1989; Roschelle, 1994; Roschelle & Goldman, 1991).

The researcher may now transcribe verbal and nonverbal activity in segments of particular interest. Transcription takes place considerably later in interaction analysis than it does in cognitive modeling or behavioral hypothesis-testing, and a much smaller proportion of the data is transcribed. Verbal transcription usually uses Jeffersonian notation (the standard for conversation analysis; Schenkein, 1978), which usually results in much higher AT:ST ratios than the simpler notation needed for behavioral and cognitive ESDA techniques. Therefore, usually only the most relevant data for the final statements are fully transcribed; a conversion of a constrained part of the data is achieved. Many ESDA software tools support text editing, but no "smart editing" environments yet exist to aid Jeffersonian transcription. Information about speech overlaps, pauses, turn durations, and interruptions can be easy to extract if separate digitized speech streams are recorded for each participant (Sellen, 1992), but the instrumentation needed to collect such data can be intrusive. Nonverbal transcription poses further challenges. Notations have been developed by Frohlich, Drew, and Monk (1994) for MKEs, cursor movements, and screen activity and by Heath and Luff (1992b) for body position and gesture when humans interact in media spaces. It is important that these notations be readable, because transcriptions play a far more important role in the naturalistic social tradition for verifying assertions and communicating research results than in the other traditions. To date, hardly any ESDA tools support entry of graphical notations (but see Roschelle & Goldman, 1991).

Finally, Statements are generalizations, interpretations, or theoretical inferences based on the data. The arrows show not only that the Body of Theory influences how statements are inferred from the data but also that these inferences influence the theories themselves, promoting their cumulative development (Jordan & Henderson, 1995). Figure 8 also reflects the tendency for interpretation to occur continually throughout analysis (rather than "popping out" just at the end), for coding to be the result rather than the means of interpretation, and for theory to develop both within and across research programs.

In summary, interaction analysis differs qualitatively from the previous two ESDA examples. There is seldom a clear-cut a priori hypothesis to test, there are fewer "obligatory" transformations of the data, and, when formally describing their work, practitioners often include what other traditions would call preliminary structuring and exploratory activities and products. Interpretation often takes place strongly throughout analysis, rather than being withheld until the final stages. Even so, interaction analysts take pains to adequately justify their statements with data. Jordan and Henderson (1995) emphasize that it must always be possible to verify or disconfirm speculations made during analysis with the data at hand and that final statements must be fully accountable to those data.

HCI and the Social Tradition of ESDA

Example in This Special Issue. Frohlich et al. (1994) characterize their own work as "an early example of work in interaction analysis," noting its strong relation to the kind of interaction analysis just described. Specifically, they show that conversation analysis can clarify how users recognize and repair "trouble" that emerges when they interact with software. In earlier work on human–human interaction (HHI), Schegloff (1991) and Schegloff, Jefferson, and Sacks (1977) identified several fundamental types of conversational repair, which Frohlich et al. now extend to HCI. Frohlich et al.'s transcripts indicate that some of Schegloff's types fit the HCI case well but that trouble is worsened by the lack of natural mutual intelligibility between human and computer. They argue that repair is the principal way that humans learn the "action grammar" of a computer system and that it is central to how work actually–rather than ideally–gets done.

Frohlich et al.'s (1994) study is firmly in the naturalistic social tradition of ESDA. The focus is on HCI as an inherently social activity, and the data are chosen for their representativeness using Thematic Sampling, rather than through conventional statistical sampling procedures. Moreover, the thesis of their article is advanced by systematically using a body of theory to interpret data, rather than through computational methods. Most important, Frohlich et al.'s work extends current microsociological theories and methods to HCI and overcomes important challenges. For example, they have developed a notation for screen activity and user actions that runs alongside the conventional verbal transcription. This allows the three activity streams to be represented in a temporally coordinated way, making it easier to detect sequential relations and to infer the "HCI order."

Further Use of the Social Tradition of ESDA in HCI. The social tradition of ESDA contributes substantively to HCI in two ways. First, many researchers have viewed HCI as a dialogue between two partners and have argued that we can fruitfully apply concepts from HHI to HCI (Frohlich et al., 1994; Frohlich & Luff, 1990; Norman & Thomas, 1990). Second, much current computing and communications technology puts humans in contact with one another, preserving some aspects of HHI but distorting others (Greenberg, 1991; Schmidt & Bannon, 1992). In both cases, the goal has been to identify processes underlying normal social interaction and to design systems that preserve the needed social "ecology" between user and artifact.

In a useful review, May (1991) outlined sociolinguistic influences on how these two types of HCI research are being conducted. May discussed the impact of conversation analysis, cognitive anthropology (Casson, 1983), speech-act theory (Winograd & Flores, 1986), and the ethnography of speaking (Linde, 1988) on issues such as collaborative work, videoconferencing, human–computer dialogues, and recovery from error. May also noted that interaction analysis, grounded theory, symbolic interactionism, and general ethnography are being used effectively to study HCI. All these approaches are highly empirical, and all emphasize observation of the "natural" situation.

Beyond its theoretical contributions, the social tradition of ESDA offers qualitative techniques that can be particularly powerful in the initial stages of any inquiry, regardless of tradition. For example, a so-called ethnographic stage often precedes the formal stages of behaviorally or cognitively oriented ESDA. Researchers often need the skills of a trained ethnographer to gain access to a site, to enlist the cooperation of participants, to choose settings, to decide on sampling strategies and events to record, and to respond effectively to unexpected turns of events. Moreover, "needs finding" is a common first step in system development and design. It involves identifying a population of users and their characteristics, work requirements, environmental considerations, and so forth. It is typically conducted in the field using observational or interview techniques and can be viewed as hypothesis-generating activity. Many recent treatments of ethnography and participant observation offer guidance here (Fetterman, 1989; Jorgensen, 1989; Miles & Huberman, 1994; chaps. in Denzin & Lincoln, 1994). Strauss's (1987) treatment of grounded theory–most notably the questions and answers at the end of the book–is particularly helpful. Although not always specialized for sequential observational data, qualitative techniques help investigators pose and refine meaningful questions, generate and maintain descriptive schemes, select data appropriately, and generally judge an investigation's progress.

Frequency Band of ESDA Event Types for the Social Tradition. A concern with the social organization of computer-supported work takes us rightward on the ESDA spectrum, where we are concerned with effective human collaboration in meetings, operations, and projects. Answers to many questions may lie in low-frequency band patterns of interaction that are more typically the province of organizational theory (see Jirotka, Gilbert, & Luff, 1992). At the low-frequency end of the ESDA spectrum, "observing" often turns into "becoming a participant in a culture." Overall, it seems that socially and organizationally challenging systems are most effectively developed if researchers are also users, users are also developers, and developers are also researchers. Despite this apparent emphasis on event types in the lower frequency band, high-frequency microanalyses can be just as helpful in determining how to support social aspects of HCI, as we saw with Frohlich et al.'s (1994) work. As further examples, Heath and Luff (1992a, 1992b) and Tang (1991) have analyzed hand gestures in meetings and operations environments and have identified the aspects of gesturing that should be transmitted if shades of meaning and the natural momentum of social interaction are to be preserved.

Boundaries of the Social Tradition

Because rigorous theory *development* is often valued more highly in the naturalistic social tradition than theory *verification* through hypothesis testing, for some practitioners the distinction between exploration and confirmation is nonexistent–or indeterminate at best. Many of the activities we have described here would be considered exploratory in other traditions. As we have noted, this makes the social tradition a rich source of techniques for identifying issues and shaping an investigation. Nonetheless, hypothesis testing can play a central role, and many practitioners call for greater efforts to test and verify theories rigorously. Tools for expressing and testing qualitative hypotheses have been built into some qualitative data analysis tools (Hesse-Biber, Dupuis, & Kinder, 1991; Huber & Garcia, 1991). Moreover, many social researchers recently have noted that little is known about how to move effectively from observation-based research findings to HCI design recommendations, and they have been trying to develop ways of doing this (Hughes, Randall, & Shapiro, 1992; Jirotka et al., 1992). The success of these efforts will determine the impact of the social tradition of ESDA on HCI research and design.

6. ESDA PROBLEMS AND ADAPTATIONS

Although ESDA studies can usually be placed within an ESDA tradition, they are seldom "pure" implementations of a tradition. We see this in G. M. Olson et al.'s (1994) work, which combines statistical and grammatical approaches to building hierarchical models of group interactions. An individual ESDA study represents very specific answers to the four questions of the generic ESDA process: "What is the issue at hand?" "What should be observed?" "What operations should be done?" "What's an acceptable type of answer?" (see Figures 1 and 5). Thus, even just implementing an existing ESDA technique requires some degree of adaptation. Occasionally, however, a new question, setting, condition, or audience will pose such a strong challenge to the usual technique that major adaptations are required. For example, the time pressures of commercial software development usually make detailed cognitive modeling impracticable (Bellotti, 1990). Moreover, designing information systems for use by highly trained experts in a fast-moving operational environment usually necessitates moving from laboratory-based to field-based ESDA settings, with all the associated complexities and uncertainties (Woods, 1993). In the face of such challenges, the most common dangers are not seeing the need or opportunity for adaptation and persevering with an inappropriate ESDA technique ("methodological fixation"), seeing the need but not knowing what to do about it ("methodological paralysis"), and seeing the need but overreacting to it ("methodological vagabonding"). In such cases, a different answer is needed to at least one of the four generic ESDA questions in Figure 1.

Figure 9 presents common ESDA problems and some possible solutions to them that involve methodological or conceptual adaptation. The columns represent 10 common problems that have been divided into three classes: failures of expertise, the threat of unacceptable AT:ST ratios, and data-based problems. The rows represent the seven ESDA dimensions from Figure 5, with the analytic operations broken out into the Eight Cs. The cell entries are potential solutions, expressed in general terms. In some cases, minor adaptations can ease ESDA problems by more effective use of the Eight Cs with the data available. In other cases, major adaptations may be necessary. The solution may require new data and may even require a researcher to step outside the boundaries of his or her tradition. In the latter case, Figure 9 should be read with Figure 5, which outlines the responses of the different traditions to the seven ESDA dimensions and thereby indicates the major adaptational routes a researcher might consider. In the most extreme cases, ESDA itself may prove to be unsuitable, as we see for cells marked "query use of ESDA." In the sections that follow, we briefly survey the 10 ESDA problems and their possible solutions and discuss some of the general issues these problems raise.

6.1. Failures of Expertise

Problem: Insufficient Domain Knowledge. In much HCI research or design work, the system or process being evaluated is well known to the researcher. However, when HCI studies are directed at highly technical or inaccessible domains of which researchers have little a priori knowledge, there is the danger that researchers will focus on superficial but easily observed behaviors. This focus leads to the generation and testing of hypotheses that are ultimately uninformative–analysis becomes a manipulation of superficials. Researchers from all traditions emphasize the importance of the preliminary "ethnographic" stage of ESDA (Jordan & Henderson, 1995; Vortac et al., 1994; Woods, 1993), during which the researcher learns about the domain from the practitioner's point of view and learns what should be recorded (see Figure 9). In some HCI domains, the researcher must acquire a significant level of formal technical expertise as well as learning local routines, customs, and practices. Additionally, domain experts–often the very people observed–can become partners in analysis.

Problem: Statement–Technique Mismatch. Without experience, it might be difficult to judge whether an ESDA technique will provide results that support the kind of Statements a researcher wants to make. After the question has been identified and a source of rigor chosen for statements, the family of ESDA techniques that allow the available data to speak to the question can be identified (see Figure 5) and considered for their feasibility.

Problem: Inadequate Experience With ESDA Technique. Expertise in an ESDA technique consists of formally acquired knowledge of the underlying ESDA tradition, technical competence in instrumentation and data collection, a feeling for the acceptable envelope of adaptation for the

Figure 9. ESDA adaptations.

	FAILURES OF EXPERTISE			AT:ST THREATS	
	Insufficient Domain Knowledge	Statement-Technique Mismatch	Inadequate Experience with ESDA Technique	Insufficient Time for Analysis	Over-whelmingly rich data
INVESTIGATIVE APPROACH	Collaborate with DE	Reexamine R&D Issues	Query use of ESDA	Query use of ESDA	Emphasize exploration
SETTING	Get familiar with setting				Find simpler setting
SAMPLING	Ask DE	Sample differently	Study and apprentice	Sample selectively	Sample selectively
FOCUS OF ANALYSIS	Ask DE	Change focus as needed	Study and apprentice		
CODE/DESCRIBE PROCESS	Collaborate with DE	Avoid certain coding styles	Study and apprentice	Avoid certain coding styles	Seek ESDA software help
ANALYTIC OPERATIONS	Collaborate with DE	Explore new subset	Practice	Seek ESDA software help	Explore new subset
chunks				Seek global structure	Seek global structure
comments			Build general familiarity	Build general familiarity	Note themes emerging
codes	Ask DE			Avoid unless value proven	Postpone
connections				Map global threads	Map global threads
comparisons	Compare with sample cases				
constraints	Select subthemes			Select subthemes	Select subthemes
conversions					Reveal hidden structure
computations				Seek ESDA software help	Summarize
SOURCES OF RIGOR	Review what is acceptable	Review what is acceptable	Study and apprentice	Review what is acceptable	

technique, and experience with inferring the scientific or design significance of results. Even if there is a good match of a chosen technique, the research or design question, and the data collected, a group of researchers may not have the expertise to implement the technique. If they try it but find the results unilluminating, then they must decide whether they implemented it inappropriately or whether there was not such a good match after all. Falsely concluding that the implementation was wrong encour-

DATA-BASED PROBLEMS

Data Type-Technique Mismatch	Statistical Short-comings	Shift in Sequence Uniqueness	Need for FB Shift	Impover-ished Data
Reexamine R&D issues	Reexamine R&D issues	Reexamine R&D issues		Query use of ESDA
Find setting yielding DT				Find richer setting
Retrieve DT from records	Sample differently	Sample differently	Sample at new FB	Sample differently
Change focus as needed	Adjust focus or statistics	Change focus as needed	Change focus as needed	Limit focus to match data
	Review for suitability	Review for suitability	Find new language	Collaborate with DE
Explore new subset	Use to test assumptions	Explore new subset	Seek ESDA software help	Explore new subset
			Explore new structurings	
		US: Analyze thematically	Note themes emerging	
	Examine set properties	RS: Probe consistencies	Combine or rewrite	
		US: Describe structure	Find new threads	
	Assess coder reliability	Vary for uniqueness		
	Aid coder reliability		Select subthemes	
	Examine data properties	US: Seek structure	Reveal hidden structure	
	Find best tests	RS: Find best tests	Reveal hidden structure	
Review what is acceptable	Abandon statistical rigor	Review what is acceptable		Review what is acceptable

KEY

R&D	Research and Design
DE	Domain Expert
FB	Frequency Band
DT	Data Type
US	Unique data sequence
RS	Replaceable data sequence

ages researchers to waste time attempting better implementations. In frustration, many researchers have developed ESDA adaptations of their own by studying current techniques, apprenticing themselves to experts, and simply trying things out in consultation with experts. B. Jordan (personal communication, October 11, 1993) has reported cases of teams of designers who have successfully used these tactics to develop their own variants of interaction analysis.

6.2. Time Management: AT:ST Threats

Problem: Insufficient Time for Analysis. Too often, researchers cannot complete an analysis with their preferred ESDA technique by the time the answer is needed. The problem may lie in the projected AT given the AT:ST ratio or in the ST itself. Figure 9 indicates possible solutions. Clearly, ST can be reduced by sampling, constraining, or changing the frequency band of analysis. AT can be reduced either by adopting a more qualitative, annotation-based technique or by using ESDA software that speeds data access, removes the need for transcription, aids data management, provides data visualization, and performs calculations.

Problem: Overwhelmingly Rich Data. If researchers pursue the interconnections among many different types of events, they will raise many more interesting HCI questions than there is time to analyze. Methodological paralysis or methodological vagabonding may set in. Solutions to the problem of overwhelmingly rich data focus on ways either of reducing the complexity of the data at the source or of extracting structure from the data after collection (Figure 9). ESDA smoothing operations such as connecting, constraining, and converting can also help the analyst discern what might be worth pursuing and where there are opportunities for computation. Alternatively, a strong-minded researcher may avoid collecting such rich data unless there is a well-formulated question and a firm notion of how the data will be analyzed (see discussion in Woods & Sarter, 1993). Paradoxically, ESDA software can both exacerbate and alleviate this problem, as we see later.

AT:ST Ratios and ESDA Traditions

The time spent applying an ESDA technique arises from different sources in different traditions, as Figures 2, 6, 7, and 8 suggest. However, certain transformations always have high AT:ST ratios, and, if a technique relies heavily on them, then the overall AT will be high. Overall, the best course is to be suspicious of ESDA activities that are known to lead to high AT:ST ratios unless their relation to the research or design question is clear.

First, detailed transcription can absorb a great deal of time. A trained typist can produce an accurate verbal-protocol transcript at an AT:ST ratio of 8:1 (Swarts, Flower, & Hayes, 1984). For interaction analysis, a rough transcript can be produced at 3:1, but more detailed transcriptions can run to 10:1 or 20:1 (Jordan & Henderson, 1995) or even to 50:1 or 100:1 in complex cases (Roschelle & Goldman, 1991). Clearly, if a researcher transcribes fairly late in analysis–when it is clear which dimensions and episodes of the data are critical to the purposes of the study and which are not–then less time is wasted. This policy is often adopted by researchers using naturalistic social techniques, but it also makes sense whenever it in unclear how data will be used to answer a question. Transcription can be partly avoided with ESDA software tools that support digitized audio storage and replay (see Sanderson, 1994).

Second, sampling and encoding policies can affect AT:ST ratios. In the behavioral tradition, Statements require an adequate sample of raw sequences, so a researcher must analyze all Logs and Recordings collected under the current sampling policy. Exhaustive rule-based linear coding may be used–with a relatively high but predictable AT:ST ratio–but calculations may involve an AT:ST ratio of much less than 1:1. In contrast, researchers in the naturalistic social tradition tend to adopt a survey-based nonlinear style of analysis, making multiple passes through the data at relatively low AT:ST ratios. Roschelle and Goldman (1991) called this *progressive refinement,* in which each successive pass exposes new features, and the most detailed analyses are reserved for the most important parts of the data.

Benefits and Traps of Software and Hardware Support

Figure 9 indicates that ESDA software can help when there are unacceptable AT:ST ratios or, as we see later, data-based mismatches. Two important assumptions behind current ESDA software development efforts are, first, that computer support will help reduce AT:ST ratios and, second, that the quality of analyses will improve. However, the AT:ST ratios that result are not always smaller because factors increasing and reducing the AT:ST ratio are in competition. Moreover, analytic quality does not necessarily increase with the use of ESDA software tools.

ESDA software and hardware can help to reduce AT:ST ratios by supporting an analyst's memory, attention, and reasoning capacity:

- Better hardware, software, and multimedia support reduce the simple access time to raw Logs and Recordings. Laser disks and CD–ROM allow researchers almost instantaneous access using time codes or more sophisticated queries and probes (constraints). These capabilities allow a researcher to ask more–and better–questions in a given amount of time and to store the results.
- Data can be transformed, or converted, from formats that require capacity-greedy human reasoning to formats that exploit human visual or auditory pattern recognition, helping researchers detect structure at different frequencies and within different event types.

- Analytical thinking can be accelerated by allowing researchers to access the products of prior thinking–such as comments, codes, and conversions–and by offering quick manipulation of these products.
- Routine calculations are simplified, and automatic inferencing and semi-automated analysis become possible on the basis of the raw data records. Computation is made practical.

Many HCI researchers using ESDA have found that computational support helps. This is especially well documented in the area of usability testing (Hoeim & Sullivan, 1994). Hammontree (cited in Weiler, 1993) developed an ESDA software tool that has reduced usability-testing AT:ST ratios from 6:1 to as low as 1:1. Hoeim (in Weiler, 1993) also reported a better rate of analysis with the Observer–Tracker–Reviewer suite and less "data panic." On the research side, Ritter and Larkin (1994) report that SMT speeded analysis of the Browser–Soar data by a factor of 10. They also note that, because SMT helped them analyze their data in minutes, rather than hours or days, a "cognitive momentum" emerged that helped them detect patterns that might otherwise have been overlooked or forgotten. K. N. Dunbar (personal communication, November 4, 1993) reported similar qualitative benefits with MacSHAPA (Sanderson et al., in press). In these cases, ESDA software tools help by "composing" the Eight Cs (see Section 6.3).

On the other hand, ESDA software can also increase AT:ST ratios and compromise analytic quality. Automated logging and mass storage not only vastly increase the ST stored, but they also alter researchers' ambitions:

- Mass storage preserves the details of observed events (Raw Sequences) and makes them permanently available for an unlimited amount of analysis.
- Raw sequences can be recorded at a finer grain than before. Replay at different speeds is possible–which effectively makes the recording grain even finer. Thus, given the same ST, there are literally more data that can be analyzed.
- It becomes tempting to record new event types simply because they can be captured easily (e.g., eye movements, gestures, routine telephone conversations). This opens the possibility of building richer and more valid theories, but the greater interconnectedness increases the AT:ST ratio.
- Because computer support makes it possible, analysts may feel compelled to study a larger quantity of data at a fine grain even when a coarser grain of analysis is more appropriate for the question.
- With a greater variety of data, the number of degrees of freedom for modeling data is vastly increased–there are more options to consider.

- ESDA software offering a full range of analysis options increases the chances that time will be wasted on unprofitable analyses. However, poorly selected ESDA software offering a limited range of analyses may channel the analyst into an inappropriate approach.
- Rather than the analyst, the data are now the ultimate authority. Analysts can find it harder to settle on an approach because of the greater odds of being proved wrong by their own data. In extreme cases, this amounts to "instant replay paralysis."

In summary, the conceptual density of analysis and the benefits and traps of ESDA software will compete to determine the AT:ST ratio for a particular analysis. ESDA software may result in an AT:ST ratio that has the researcher spending just as much time analyzing data as before. The critical criterion for judging ESDA software is not whether AT:ST ratios are higher or lower but, instead, whether HCI design implications are seen more clearly and whether theories of HCI become richer, more interesting, and more powerful. ESDA software does not guarantee greater theoretical richness; theoretical richness depends instead on how closely the researcher knows the data and whether good questions are being asked. However, if ESDA software supports the transformations the researcher needs, then quality should increase.

As Sanderson (1994) has discussed in detail, no single ESDA software application can serve a researcher's entire needs across a variety of research and design questions. ESDA software should be chosen on the basis of how useful its strengths are and how readily its limitations can be patched by other software. Much excellent ESDA work has been done by combining the capabilities of reliable productivity software such as spreadsheets, database programs, statistical packages, and drawing programs (Bainbridge & Sanderson, in press; Flor & Hutchins, 1991; Hoeim & Sullivan, 1994; Richards & Richards, 1994).

6.3. Data-Based Problems

Problem: Data Type–Technique Mismatches. If a researcher chooses an ESDA technique because of a commitment to a particular theoretical orientation, then the event types collected must make sense for that technique. For example, verbal protocols should not include any speculation by the speaker about cognitive processes, and Logs and Recordings used in interaction analysis should include auditory information on intonation and timing and visual information on body orientation and gesture. When there is a mismatch between the type of data available and the preferred ESDA technique, researchers are often faced with major adaptations.

Event types themselves sometimes seem to suggest techniques that, on closer examination, are inappropriate. For example, if recordings of natural talk are available, researchers with cognitive backgrounds may gravitate toward verbal protocol analysis. However, participants in natural exchanges achieve a great deal through sociolinguistic pragmatics that are simply not detectable in the propositional content of utterances (Hutchins & Klausen, in press). An ESDA technique is needed in which communicative and interaction phenomena can be expressed (Clark & Brennan, 1991; Frohlich et al., 1994; Heath & Luff, 1992a; Jordan & Henderson, 1995).

Nonetheless, particular event types are not tied inexorably to particular ESDA techniques. For example, gesture is a high-frequency event type, but, in many contexts, it can be interpreted only with the aid of theories of human communication and an understanding of human purposes at the project level (Heath & Luff, 1992a, 1992b; Tang, 1991). The theoretical context rather than the event type determines the technique.

Problem: Statistical Shortcomings Statistical procedures can be used for exploratory or confirmatory purposes, but the results will be meaningless if raw sequences cannot be encoded reliably, if they have too few examples of the events of interest, if the temporal relations are too noisy to be detected by the algorithm chosen, if data properties violate underlying statistical models, or if sequential dependencies change within the sequence. For example, both G. M. Olson et al. (1994) and Vortac et al. (1994) ensured their data were not sparse by collapsing coding categories whenever appropriate before running statistical tests. It can be difficult to distinguish independent and dependent variables whenever the events that people react to are partly the result of their own prior actions (Laws & Barber, 1989). Moreover, it may be impossible to calculate error of estimate and chance occurrence in event sequences that are not fully under experimental control (Simon & Kaplan, 1989). Last, data used for confirmatory analyses must be separated from data used for exploration. If there is a limited amount of data or if events of interest are so rare that they cannot be sampled, researchers need to find alternative statistics (nonparametric, descriptive) or abandon statistical rigor entirely.

Problem: Shift in Sequence Uniqueness. In some cases, the events recorded are "unique" because they are unlikely ever to happen again in the same form (consider critical incidents, errors, discoveries, or misunderstandings). In other cases, events are "repeated" or "replaceable" because they are just as likely to be found in one data sequence as in another. When data consist of unique sequences, techniques supporting the reliable and valid interpretation of single episodes are needed (Hutchins & Klausen, in press). When data consist of repeated sequences, formal replications and empirical contrasts become possible. Figure 9 shows that shifts between unique and replaceable data sequences principally affect how the Eight Cs are used, but they can also lead to the adoption of different sources of rigor for Statements such as narrative or model-based versus statistical.

Problem: Need for Frequency Band Shift. If a researcher has always paid close attention to regularities at one band of frequencies in the data, then, in a new setting or under new conditions, important regularities at other frequencies may be missed. Much exploration represents a search for a way to describe events, and coding can be a bandpass filter as well as a filter for different possible meanings and theories. Constraining, coding, and converting data into new visual forms can help overcome entrapment within an inappropriate frequency band and event type (for examples, see Hoeim & Sullivan, 1994; G. M. Olson et al., 1994; Sanderson, 1994; J. B. Smith et al. 1993).

Problem: Impoverished Data. An ESDA technique cannot be used if logs and recordings do not capture aspects of raw sequences that the technique requires, such as specific data types, particular event structures, or enough time to detect low-frequency periodicities. For example, in a cooperative work environment, the activities and utterances of important off-site players may be unavailable, electronic communications may not be recorded, interactions with devices may not be logged, and important phases of work may be overlooked. If the data are also "unique," or if the researcher cannot return to the field, then domain experts or the participants themselves might fill in some gaps with post hoc commentaries. Ultimately, a non–ESDA approach to data collection–such as simulations, interviews, critical incident reports, examination of artifacts, and so forth–should be considered.

Promoting Adaptation by Composing the Eight Cs

In showing the many different paths between Logs and Recordings and Statements, Figures 6, 7, and 8 exhibit two noteworthy features. Each uses at least seven of the Eight Cs, and many transitions among nodes involve two or more of the Eight Cs. When developing coding categories, analysts might perform a preliminary coding, constrain their current focus to a subset of these codes, and compare members of this subset to see if the coding is coherent. The solutions suggested in Figure 9 often involve using a new subset of the Eight Cs, and this can lead to "composing" the Eight Cs in powerful ways. Instead of coding and calculating, a researcher might resort to commenting, constraining on the basis of some property of the comments, and then converting the result to a graphical form. ESDA software environments should support fundamental transformations like the Eight Cs and allow them to be composed. Multiple active software links should connect the products of these transformations so that the researcher can quickly move between different data views and contexts (Sanderson, 1994; Sanderson et al., in press).

Being able to compose the Eight Cs liberally–examining and comparing results while always remaining connected to other products as well as to the raw data–opens a vast space for exploring data and adapting techniques and is a simple extension of Tukey's (1977) EDA philosophy to sequential data. Composition may not only speed judgments, but it can also–if well managed intellectually–exercise a configural effect on the analyst's ability to make inferences. Relations may appear and judgments may become possible that would not otherwise have–and theory becomes richer and better substantiated as a result.

Qualitative Impact of ESDA Software

Throughout this article, we have identified some of the factors that can promote usefulness and theoretical richness when ESDA is applied to HCI problems. The preceding comments about software support highlight the fact that ESDA is a task for HCI as well as a methodology for it, just as Waterman and Newell (1971, 1972) proposed that protocol analysis was a task for artificial intelligence as well as a methodology for it. Our view of how ESDA should be supported is consistent with current thinking in cognitive engineering about how the inferential or "knowledge-based" reasoning of human operators might be aided by providing them with information displays that reveal the invariant structure of a complex system (Bennett & Flach, 1992; Rasmussen, Pejtersen, & Goodstein, in press; Vicente & Rasmussen, 1992; Woods, 1993). Effective displays allow normal and abnormal workings to be directly perceived rather than having to be laboriously inferred. For ESDA, we are grappling with the inverse of this problem. Our goal is to provide analysts with a toolbox of generic transformational operations such as the Eight Cs–which, when accompanied by a theoretical outlook or tradition-based framework, will help analysts actively discover the invariant structure (or structures) in a set of observational data.

If ESDA software offers a "generative" environment in which fundamental analytic operations can be composed in vastly different ways in order to forge new techniques and support different traditions, then it also offers a medium in which ESDA practitioners from different traditions can work cooperatively and communicate. For example, a cognitive engineer and a conversation analyst may use the same collection of computational tools to perform their respective analyses. Not only can they share the same body of data, but they have more direct access to the results of each other's analyses. Most important, they can inspect the evidence and the physical remnants of the process by which each other's results were achieved. Thus, by thinking of ESDA as the fluid composition for different purposes of a set of analytic operations, rather than as a collection of unconnected techniques, interdisciplinary collaboration on HCI projects can be supported (Jones, Kaplan, Sanderson, & Star, 1994).

6.4. Challenges of Video and Reporting ESDA Studies

In HCI, one increasingly hears the term *video analysis* for video-intense ESDA. Elsewhere, we have raised objections that this describes an ESDA technique no more than *telescope analysis* describes astronomy (Fisher & Sanderson, 1993) and that it does not point to any specific methods for drawing scientific or design-related inferences from video. We agree with Carter and B. Anderson (1989) that video analysis should be "as firmly grounded in an adequate methodology (broadly construed) as any other investigative strategy" (p. 113). Carter and B. Anderson offer an excellent discussion of how AT:ST ratios can become unnecessarily inflated by video.

Video supports many ESDA techniques, but it is probably most pervasively and creatively used within the social tradition, in which many researchers view video as a medium of analysis in its own right, rather than merely as a high-resolution storage repository for raw sequences until they are transformed into "analyzable" codes and numbers (Roschelle & Goldman, 1991). For example, qualitative reasoning can be aided by placing annotations directly on video images (van der Vlugt, Kruk, van Erp, & Geuze, 1992), by having video itself trigger annotations or other products associated with interpretation as it replays (Hesse-Biber et al., 1991; Roschelle & Goldman, 1991), by being able to carry out a hierarchical search through a video source to specific parts while remaining in touch with the larger temporal context of the whole source (Mills, Cohen, & Wong, 1992), or by splicing together video segments into a document that can be used to make comparisons, refine theory, or make presentations. These capabilities reflect an emphasis on staying grounded in raw data that is stronger in the social tradition than in the behavioral and cognitive traditions and that is shared with EDA (Tukey, 1977).

However, these capabilities also raise questions about how colleagues might be able to assess statements derived from video-based ESDA without reviewing an entire analysis. Even when video is not used, critics often complain that ESDA studies simply report conclusions and that it is difficult to see how the conclusions were drawn from the evidence. In ESDA studies, the evidence is usually extremely bulky, and there is no accepted standard for reporting. It helps to publish ESDA products after a certain number of transformations have brought them to a form that speaks clearly to the issue at hand. An example is Roth et al.'s (1987) article on the effectiveness of dialogues between humans and computer advisors in trouble-shooting situations, which presents the "protocols" of their full sample of subjects–or, more precisely, the chunked and connected low-frequency band codings of these protocols.

This problem becomes more acute when the underlying evidence cannot readily be represented on paper, either because it is dynamic audiovisual information (facial expression, timing, gesture, human–computer dialogue with various types of input devices) or because it consists of networks of inferences represented in hypermedia that have to be traversed in time in order to be understood. In HCI design, tapes of highlights are often created; in research contexts, however, evidence of this kind currently can be shown readily only at scientific meetings and may not be considered to be rigorous evidence. Electronic documents with video and audio inserts can be created and shared, but, until multimedia scientific journals come fully into the mainstream, such documents will not become an institutionalized form of communicating high-quality ESDA results. In the meantime, frame grabs accompanied by transcription can summarize and simulate some contact with the original data.

7. CONCLUSIONS

Observational studies have always been a part of HCI, particularly in the context of usability studies. However, current HCI questions that do not readily lend themselves to modeling or to controlled laboratory experimentation will lead us even more to observational studies. To successfully answer such questions, we need to understand how to tap the richness of observational studies more efficiently, either by adopting well-defined ESDA techniques used by certain communities of practice or by adapting existing ESDA techniques to make a customized technique. Adoptions require a thorough understanding of the intellectual context of the technique being borrowed. Adaptations require not only knowledge of the intellectual contexts of several techniques but an appreciation of whether–and how–they combine coherently. Moreover, if adaptations are to be successfully supported by ESDA software, such software must offer general operations that can be used, omitted, and combined in many different ways. Already we see advances in these directions. Hand in hand with such technical developments must be conceptual developments, or a unified way of thinking about how sequential data can be explored–cutting across many traditions–that provides a framework for HCI investigators who have chosen ESDA as their tool. In this article, we have taken the first steps on this path.

NOTES

Acknowledgments. We thank Neville Moray, Peter Polson, Gary Olson, Sharon Carver, and two anonymous reviewers for very helpful comments on earlier forms of this article.

Support. This research was supported by NASA–Ames Research Center Grant NAG 2-731 (Sandy Hart); Wright Patterson Air Force Base (Ken Boff, Mike McNeese) through Logicon Technical Services (Lawrence Wolpert); Aeronautical and Maritime Research Laboratory, DSTO, Australia (Jeremy Manton, Simon Oldfield, Dominic Drumm); and University of Illinois Research Board.

Authors' Present Addresses. Penelope M. Sanderson, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana–Champaign, 1206 West Green Street, Urbana, IL 61801. E-mail: psanders@psych.uiuc.edu; Carolanne Fisher, MAYA Design Group, 2100 Wharton Street, Pittsburgh, PA 15219. E-mail: fisher@maya.com.

HCI Editorial Record. First manuscript received March 9, 1993. Revision received February 28, 1994. Accepted by Gary M. Olson. Final manuscript received June 14, 1994. – *Editor*

REFERENCES

Adler, P. A., & Adler, P. (1994). Observational techniques. In N. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 377–392). Thousand Oaks, CA: Sage.

Allison, P. D., & Liker, J. K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. *Psychological Bulletin, 2,* 392–403.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, R. J., Heath, C. C., Luff, P., & Moran, T. P. (1993). The social and the cognitive in human–computer interaction. *International Journal of Man–Machine Studies, 38,* 999–1016.

Bailey, W. A., & Kay, E. J. (1987). Structural analysis of verbal data. *Proceedings of CHI+GI,* 297–300. New York: ACM.

Bainbridge, L., & Sanderson, P. M. (in press). Verbal protocol analysis. In J. Wilson & E. N. Corlett (Eds.), *Evaluation of human work* (2nd ed.). London: Taylor & Francis.

Bakeman, R. (1983). Computing lag sequential statistics: The ELAG program. *Behavior Research Methods & Instrumentation, 15,* 530–535.

Bakeman, R., & Gottman, J. (1986). *Observing interaction: An introduction to sequential analysis.* Cambridge, England: Cambridge University Press.

Bales, R. F. (1950). *Interaction process analysis.* Reading, MA: Addison-Wesley.

Bellotti, V. M. E. (1990). A framework for assessing applicability of HCI techniques. In D. Diaper, D. Gilmore, G. Cockton, & B. Shackel (Eds.), *Human–Computer Interaction–INTERACT'90,* 213–218. Amsterdam: North-Holland.

Bennett, K., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34,* 513–534.

Benson, D., & Hughes, J. (1991). Method: Evidence and inference–Evidence and inference for ethnomethodology. In G. Button (Ed.), *Ethnomethodology and the human sciences* (pp. 109–136). Cambridge, England: Cambridge University Press.

Bevan, N., & Macleod, M. (1994). Usability measurement in context. *Behavior Information and Technology, 13,* 132–145.

Bhaskar, R., & Simon, H. A. (1977). Problem-solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive Science, 1,* 193–215.

Blumer, H. (1954). What is wrong with social theory? *American Sociological Review, 19,* 3–10.

Bolt, R., & Herranz, E. (1992). Two-handed gesture in multi-modal dialog. *Proceedings of the ACM Symposium on User Interface Software and Technology,* 7–14. New York: ACM.

Boose, J. H. (1984). *Expertise transfer for expert system design.* Amsterdam: Elsevier.

Bovair, S., Kieras, D., & Polson, P. (1990). The acquisition and performance of text-editing skill: A cognitive complexity analysis. *Human–Computer Interaction, 5,* 1–48.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human–computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Carter, K., & Anderson, B. (1989). Can video research escape the technology? Some reflections on the problems and possibilities of A.V. research. *SIGCHI Bulletin, 21,* 112–114.

Casson, R. W. (1983). Schemata in cognitive anthropology. *Annual Review of Anthropology, 12,* 429–462.

Cicourel, A. V. (1990). The integration of distributed knowledge in collaborative medical diagnosis. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work* (pp. 221–242). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Clancey, W. J. (1993). Situated action: A neuropsychological interpretation. Response to Vera and Simon. *Cognitive Science, 17,* 87–116.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Conti, R., Cuomo, D., & Rogers, J. C. (1994). *UTEST: Usability test support tool.* Unpublished manuscript, Mitre Corporation, Bedford, MA.

Cooke, N. J., Neville, K. J., & Rowe, A. L. (in press). Procedural network representations of sequential data. *Human–Computer Interaction.*

Denzin, N. K., & Lincoln, Y. S. (1994). *Handbook of qualitative research.* Thousand Oaks, CA: Sage.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Cambridge, MA: MIT Press.

Faraone, S. V., & Dorfman, D. D. (1987). Lag sequential analysis: Robust statistical methods. *Psychological Bulletin, 101,* 312–323.

Fetterman, D. M. (1989). *Ethnography step by step.* Newbury Park, CA: Sage.

Fisher, C. (1988). Advancing the study of programming with computer-aided protocol analysis. In G. Olson, E. Soloway, & S. Sheppard (Eds.), *Empirica studies of programmers: 1987 workshop* (pp. 198–216). Norwood, NJ: Ablex.

Fisher, C. (1991). *Protocol Analyst's Workbench: Design and evaluation of computer-aide* *protocol analysis.* Unpublished PhD thesis, Carnegie Mellon University, Department of Psychology, Pittsburgh.

Fisher, C., & Sanderson, P. M. (1993). Exploratory sequential data analysis: Traditions, techniques and tools: Report of the CHI '92 workshop. *SIGCHI Bulletin, 25,* 31–40.

Flor, N. V., & Hutchins, E. (1991). Analyzing distributed cognition in software teams: A case study of team programming during perfective software maintenance. In J. Joenemann-Belliveau, T. G. Moher, & S. P. Robertson (Eds.), *Empirical studies of programmers: Fourth workshop* (pp. 36–64). Norwood, NJ: Ablex.

Frohlich, D., Drew, P., & Monk, A. (1994). Management of repair in human–computer interaction. *Human–Computer Interaction, 9,* 385–425. [Included in this Special Issue.]

Frohlich, D., & Luff, P. (1990). Applying the technology of conversation to the technology for conversation. In P. Luff, N. Gilbert, & D. Frohlich (Eds.), *Computers and conversations* (pp. 187–220). London: Academic.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* New York: Basic.

Garfinkel, H. (1967). *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall.

Geertz, C. (1973). *The interpretation of cultures.* New York: Basic.

Gentner, D., & Stevens, A. (1983). *Mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Goldman-Segall, R. (1992). Collaborative virtual communities: Using *Learning Constellations,* a multimedia ethnographic research tool. In E. Barrett (Ed.), *Sociomedia: Multimedia, hypermedia, and the social construction of knowledge* (pp. 257–294). Cambridge, MA: MIT Press.

Gottman, J. M., & Roy, A. K. (1990). *Sequential analysis: A guide for behavioral researchers.* Cambridge, England: Cambridge University Press.

Gould, J. D. (1988). How to design usable systems. In M. Helander (Ed.), *Handbook of human–computer interaction* (pp. 757–789). Amsterdam: North-Holland.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human–Computer Interaction, 8,* 237–308.

Green, T. R. G. (1991). User modeling: The information-processing perspective. In J. Rasmussen, H. B. Andersen, & N. O. Bernsen (Eds.), *Human–computer interaction: Research directions in cognitive science. European perspectives* (Vol. 3, pp. 27–57). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Greenberg, S. (1991). *Computer-supported cooperative work and groupware.* San Diego: Academic.

Grudin, J. (1990). The computer reaches out: The historical continuity of interface design. *Proceedings of the CHI '90 Conference on Human Factors in Computer Systems,* 261–268. New York: ACM.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.

Guindon, R. (1990). A multidisciplinary perspective on dialogue structure in user–advisor dialogues. In R. Guindon (Ed.), *Cognitive science and its applications for human–computer interaction* (pp. 163–197). Hillsdale, NJ: Lawrence Erlbaum

Associates, Inc.

Harrison, B. L. (1991). Video annotation and multimedia interfaces: From theory to practice. *Proceedings of the Human Factors Society 35th Annual Meeting,* 319–323. Santa Monica, CA: HFS.

Hartson, J. R., Siochi, A. C., & Hix, D. (1990). The UAN: A user-oriented representation for direct manipulation interface designs. *ACM Transactions on Information Systems, 8,* 181–203.

Hartwig, F., & Dearing, B. E. (1979). *Exploratory data analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences). Newbury Park, CA: Sage.

Heath, C., & Luff, P. (1992a). Collaboration and control: Crisis management and multimedia technology in London Underground line control rooms. *Computer Supported Cooperative Work (CSCW), 1,* 69–94.

Heath, C., & Luff, P. (1992b). Media space and communicative asymmetries: Preliminary observations of video-mediated interaction. *Human–Computer Interaction, 7,* 315–346.

Heritage, J. (1988). Explanations as accounts: A conversation analytic perspective. In C. Antaki (Ed.), *Analyzing everyday explanation: A casebook of methods* (pp. 127–144). Beverly Hills, CA: Sage.

Hesse-Biber, S., Dupuis, P., & Kinder, T. S (1991). HYPEResearch: A computer program for the analysis of qualitative data with an emphasis on multimedia analysis and hypothesis testing. *Social Science Computer Review, 9,* 452–460.

Hoc, J. M. (1986). *The psychology of planning.* London: Academic.

Hoeim, D., & Sullivan, K. (1994). Designing and using integrated data collection and analysis tools: Challenges and considerations. *Behavior and Information Technology, 13,* 160–170.

Hollnagel, E. (1979). *A framework for the description of operator behavior* (Technical Report N-35-79). Roskilde, Denmark: Risø National Laboratory, Electronics Department.

Holstein, J. A., & Gubrium, J. F. (1994). Phenomenology, ethnomethodology, and interpretive practice. In N. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 262–272). Thousand Oaks, CA: Sage.

Hoppe, H. U., & Schiele, F. (1992). Towards task models for embedded information retrieval. *Proceedings of the CHI '92 Conference on Human Factors in Computing Systems,* 173–180. New York: ACM.

Huber, G. L., & Garcia, C. M. (1991). Computer assistance for testing hypotheses about qualitative data: The software package AQUAD 3.0. *Qualitative Sociology, 14,* 325–348.

Hughes, J. A., Randall, D., & Shapiro, D. (1992). Faltering from ethnography to design. *Proceedings of the CSCW '92 Conference on Computer-Supported Cooperative Work,* 115–122. New York: ACM.

Hutchins, E. (1991). The social organization of distributed cognition. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 283–307). Washington, DC: American Psychological Association.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge, MA: MIT Press.

Hutchins, E., & Klausen, T. (in press). Distributed cognition in an airline cockpit. In D. Middleton & Y. Engestrom (Eds.), *Cognition and communication at work.*

Cambridge, England: Cambridge University Press.

Irving, S., Polson, P., & Irving, J. E. (1994). A GOMS analysis of the advanced automated cockpit. *Proceedings of the CHI '94 Conference on Human Factors in Computer Systems*, 344–350. New York: ACM.

Jirotka, M., Gilbert, N., & Luff, P. (1992). On the social organization of organizations. *Computer Supported Cooperative Work (CSCW), 1*, 95–118.

John, B. (1994). *A database for analyzing sequential behavioral data and their associated cognitive models* (Technical Report CMU-HCII-94-101). Pittsburgh: Carnegie Mellon University, School of Computer Science, Human–Computer Interaction Institute.

Jones, P., Kaplan, S., Sanderson, P. M., & Star, L. S. (1994). Alchemist: Support for emergent models of work practices in collaborative systems. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 373–378. New York: IEEE.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences, 4*, 39–103.

Jorgensen, D. L. (1989). *Participant observation: A methodology for human studies.* Newbury Park, CA: Sage.

Karat, J. (1988). Software evaluation methodologies. In M. Helander (Ed.), *Handbook of human–computer interaction* (pp. 891–903). Amsterdam: North-Holland.

Kendon, A. (1982). The organization of behavior in a face-to-face interaction: Observations on the development of a methodology. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 440–505). Cambridge, England: Cambridge University Press.

Kieras, D., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8*, 255–273.

Kieras, D., & Polson, P. (1985). An approach to the formal analysis of user complexity. *International Journal of Man–Machine Studies, 22*, 365–394.

Kirwan, A., & Ainsworth, J. (1992). *A guide to task analysis.* London: Taylor & Francis.

Kowalski, B., & VanLehn, K. (1988). Cirrus: Inducing subjective models from protocol data. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, 623–629. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Laws, J., & Barber, P. (1989). Video analysis in cognitive ergonomics: A methodological perspective. *Ergonomics, 32*, 1303–1318.

Lea, M., & Spears, R. (1991). Computer-mediated communication, de-individuation and group decision making. In S. Greenberg (Ed.), *Computer-supported cooperative work and groupware* (pp. 155–173). San Diego: Academic.

Leinhardt, S., & Wasserman, S. S. (1979). Exploratory data analysis: An introduction to selected methods. *Sociological Methods, 8*, 311–365.

Linde, C. (1988). The quantitative study of communicative success: Politeness and accidents in aviation discourse. *Language in Society, 17*, 375–399.

Mackay, W. E. (1989). EVA: An experimental video annotator for symbolic analysis of video data. *SIGCHI Bulletin, 21*, 68–71.

MacWhinney, B. (1991). *The CHILDES Project: Tools for analyzing talk.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

May, L. (1991). Sociolinguistic research on human–computer interaction: A perspective from anthropology. *Social Science Computer Review, 9*, 529–540.

Mayhew, D. (1990). Cost-justifying human factors support–A framework. *Proceedings of the Human Factors Society 34th Annual Meeting,* 834–838. Santa Monica, CA: HFS.

McGrath, J. (1992). Groups interacting with technology: The complex and dynamic fit of group, task technology, and time. *Proceedings of the CSCW '92 Conference on Computer-Supported Cooperative Work,* 4. New York: ACM.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Beverly Hills, CA: Sage.

Mills, M., Cohen, J., & Wong, Y. Y. (1992). A magnifier tool for video data. *Proceedings of the CHI '92 Conference on Human Factors in Computer Systems,* 93–98. New York: ACM.

Moll van Charante, E., Cook, R. I., Woods, D. D., Yue, L., & Howie, M. B. (1993). Human–computer interaction in context: Physician interaction with automated intravenous controllers in the heart room. In H. G. Stassen (Ed.), *Analysis, design and evaluation of man–machine systems 1992* (pp. 263–274). Oxford, England: Pergamon.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A., & Card, S. (1985). The prospects for psychological science in human computer interaction. *Human–Computer Interaction, 1,* 209–242.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nielsen, J. (1993). *Usability engineering.* San Diego: Academic.

Nielsen, J. (Ed.). (1994). Special issue on usability laboratories. *Behavior and Information Technology, 13,* 3–197.

Noldus, L. P. J. J. (1991). The Observer: A software system for collection and analysis of observational data. *Behavior Research Methods, Instruments, and Computers, 23,* 415–429.

Norman, M., & Thomas, P. (1990). The very idea: Informing HCI design from conversation analysis. In P. Luff, N. Gilbert, & D. Frohlich (Eds.), *Computers and conversations* (pp. 51–65). London: Academic.

Olson, G. M., Herbsleb, J. D., & Rueter, H. H. (1994). Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques. *Human–Computer Interaction, 9,* 427–472. [Included in this Special Issue.]

Olson, G. M., & Olson, J. S. (1991). User-centered design of collaboration technology. *Journal of Organizational Computing, 1,* 61–83.

Olson, G. M., Olson, J. S., & Kraut, R. E. (1992). Introduction to this special issue on computer-supported cooperative work. *Human–Computer Interaction, 7,* 251–256.

Olson, J. S., & Olson, G. M. (1990). The growth of cognitive modeling in human–computer interaction since GOMS. *Human–Computer Interaction, 5,* 221–265.

Peck, V. A., & John, B. E. (1992). Browser–Soar: A computational model of a highly interactive task. *Proceedings of the CHI '92 Conference on Human Factors in Computer Systems,* 165–172. New York: ACM.

Pitman, K. M. (1985). *CREF: An editing facility for managing structured text* (AI Memo 829). Cambridge, MA: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Poltrock, S., & Nasr, M. G. (1989). *Protocol analysis: A tool for analyzing human–computer interaction* (MCC Technical Report ACT-HI-186-89). Austin, TX: Microelectronics and Computer Technology Corporation.

Posner, M. (1989). *Foundations of cognitive science.* Cambridge, MA: MIT Press.

Randle, J. D., & Szostak, T. K. (1993). *The Observational Coding System of Tools–A modular system integrating observational research and computer analysis.* Unpublished manuscript, Triangle Research Collaborative, Inc., Research Triangle Park, NC.

Rasmussen, J., & Jensen, A. (1973). *A study of mental procedures in electronic trouble-shooting* (Technical Report Risø-M-1582). Roskilde, Denmark: Risø National Laboratory, Electronics Department.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (in press). *Cognitive engineering: Concepts and applications.* New York: Wiley.

Richards, T. J., & Richards, L. (1994). Using computers in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 445–462). Thousand Oaks, CA: Sage.

Ritter, F. E. (1992). *A methodology and software environment for testing process models' sequential predictions with protocols.* Unpublished PhD thesis, Carnegie Mellon University, Department of Psychology, Pittsburgh.

Ritter, F. E., & Larkin, J. H. (1994). Developing process models as summaries of HCI action sequences. *Human–Computer Interaction, 9,* 345–383. [Included in this Special Issue.]

Roschelle, J. (1994). *Video analysis tools for HCI: Keep it simple, flexible, portable.* Unpublished manuscript, Institute for Research on Learning, Palo Alto, CA.

Roschelle, J., & Goldman, S. (1991). VideoNoter: A productivity tool for video data analysis. *Behavior Research Methods, Instruments, and Computers, 23,* 219–224.

Roschelle, J., Pea, R., & Trigg, R. (1990). *VideoNoter: A tool for exploratory video analysis* (Technical Report IRL 90-0021). Palo Alto, CA: Institute for Research on Learning.

Roth, E., Bennett, K., & Woods, D. D. (1987). Human interaction with an 'intelligent' machine. *International Journal of Man–Machine Studies, 27,* 479–525.

Rumelhart, D. E., McClelland, J., & PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory and Cognition, 17,* 759–769.

Sackett, G. P. (1978). *Observing behavior* (Vol. 2). Baltimore: University Park Press.

Sanderson, P. M. (1991). *ESDA: Exploratory sequential data analysis* (Technical Report EPRL-91-04). Urbana: University of Illinois at Urbana–Champaign, Department of Mechanical and Industrial Engineering, Engineering Psychology Research Laboratory.

Sanderson, P. M. (1994). *ESDA: Software* (Technical Report EPRL-94-01). Urbana: University of Illinois at Urbana–Champaign, Department of Mechanical and Industrial Engineering, Engineering Psychology Research Laboratory.

Sanderson, P. M., James, J. M., & Seidler, K. S. (1989). SHAPA: An interactive software environment for protocol analysis. *Ergonomics, 32,* 1271–1302.

Sanderson, P. M., Scott, J. J. P., Mainzer, J., Johnston, T., Watanabe, L. M., & James, J. M. (in press). MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human–Computer Studies.*

Sanderson, P. M., Verhage, A. G., & Fuld, R. B. (1989). Verbal protocol and state space approaches to continuous process control *Ergonomics, 32,* 1343–1372.

Schatzman, L., & Strauss, A. (1973). *Field research: Strategies for a natural sociology.* Englewood Cliffs, NJ: Prentice-Hall.

Schegloff, E. A. (1991). Repair after next turn: The last structurally provided defense of intersubjectivity. *American Journal of Sociology, 97,* 1295–1345.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preferences for self-correction in the organization of repair in conversation. *Language, 53,* 361–382.

Schenkein, J. (1978). Explanation of transcription notation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. xi–xvi). New York: Academic.

Scherer, K. R., & Ekman, P. (1982). Methodological issues in studying nonverbal behavior. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 1–44). Cambridge, England: Cambridge University Press.

Schmidt, K., & Bannon, L. (1992). Taking CSCW seriously: Supporting articulation work. *Computer Supported Cooperative Work (CSCW), 1,* 7–40.

Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. In N. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 118–137). Thousand Oaks, CA: Sage.

Sellen, A. (1992). Speech patterns in video-mediated conversations. *Proceedings of the CHI '92 Conference on Human Factors in Computer Systems,* 49–59. New York: ACM.

Shrager, J., & Klahr, D. (1986). Instructionless learning about a complex device: The paradigm and observations. *International Journal of Man–Machine Studies, 25,* 153–189.

Simon, H. A., & Kaplan, C. (1989). Foundations of cognitive science. In M. Posner (Ed.), *Foundations of cognitive science* (pp. 1–47). Cambridge, MA: MIT Press.

Siochi, A. C., & Hix, D. (1991). A study of computer-supported user interface evaluation using maximal repeating pattern analysis. *Proceedings of the CHI '89 Conference on Human Factors in Computer Systems,* 183–188. New York: ACM.

Smith, J. B., Smith, D. K., & Kuptsas, E. (1993). Automated protocol analysis. *Human–Computer Interaction, 8,* 101–145.

Star, S. L. (1989). The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. In L. Gasser & M. N. Huhns (Eds.), *Distributed artificial intelligence* (Vol. 2, pp. 37–54). London: Pitman.

Strauss, A. (1987). *Qualitative analysis for social scientists.* Cambridge, England: Cambridge University Press.

Suchman, L. (1987). *Plans and situated actions: The problem of human–machine communication.* New York: Cambridge University Press.

Suchman, L., & Trigg, R. (1991). Understanding practice: Video as a medium for reflection and design. In J. Greenbaum & M. Kyng (Eds.), *Design at work* (pp. 65–89). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Swarts, H., Flower, L. S., & Hayes, J. R. (1984). Designing protocol studies of the writing process: An introduction. In R. Beach & L. S. Bridwell (Eds.), *New directions in composition research* (pp. 53–71). New York: Guilford.

Sweeney, M., Maguire, M., & Shackel, B. (1993). Evaluating user–computer interaction: A framework. *International Journal of Man–Machine Studies, 38,* 689–711.

Tang, J. (1991). Findings from observational studies of collaborative work. *International Journal of Man–Machines Studies, 34,* 134–160

Tapp, J. (1994). *MOOSES: Multiple Option Observation System for Experimental Studies.* Unpublished manuscript, Vanderbilt University.

Tatar, D. (1989). Using video-based observation to shape the design of a new technology. *SIGCHI Bulletin, 21,* 108–111.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

van der Vlugt, M. J., Kruk, M. R., van Erp, A. M. M., & Geuze, R. H. (1992). CAMERA: A system for fast and reliable acquisition of multiple ethological records. *Behavior Research Methods, Instruments, and Computers, 24,* 147–149.

van Hooff, J. A. R. A. M. (1982). Categories and sequences of behavior: Methods of description and analysis. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 362–439). Cambridge, England: Cambridge University Press.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics, 22,* 589–606.

Vortac, O. U., Edwards, M. B., & Manning, C. A. (1994). Sequences of actions for individual and teams of air traffic controllers. *Human–Computer Interaction, 9,* 319–343. [Included in this Special Issue.]

Waterman, D. A., & Newell, A. (1971). Protocol analysis as a task for artificial intelligence. *Artificial Intelligence, 2,* 285–318.

Waterman, D. A., & Newell, A. (1972). *PAS-II: An interactive task-free version of an automatic protocol analysis system.* Pittsburgh: Carnegie Mellon University, Department of Computer Science.

Weedman, J. (1991). Conversational flexibility in a computer conference used in professional education. In S. Greenberg (Ed.), *Computer-supported cooperative work and groupware* (pp. 175–190). San Diego: Academic.

Weiler, P. (1993). Software for the usability lab: A sampling of current tools. *Proceedings of the InterCHI '93 Conference on Human Factors in Computer Systems,* 57–60. New York: ACM.

Whiteside, J., & Wixon, D. (1987). Discussion: Improving human–computer interaction–A quest for cognitive science. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human–computer interaction* (pp. 353–365). Cambridge, MA: MIT Press.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design.* Reading, MA: Addison-Wesley.

Woods, D. D. (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In G. Klein, R. Calderwood, & J. Orasanu (Eds.), *Decision making in action: Models and methods* (pp. 29–51). Norwood, NJ: Ablex.

Woods, D. D., & Sarter, N. (1993). Evaluating the impact of new technology on human–machine cooperation. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and validation of complex systems: Human factors issues* (pp. 133–158). Bonn: Springer-Verlag.