Using Cognitive Ethnography to Study Instruction

Robert F. Williams, Lawrence University, Appleton, WI 54912-0599, robert.f.williams@lawrence.edu

Abstract: Advances in digital media support a form of inquiry called cognitive ethnography. Cognitive ethnography employs traditional ethnographic methods to build knowledge of a community of practice and then applies this knowledge to the micro-level analysis of specific episodes of activity. The principal aim of cognitive ethnography is to reveal how cognitive activities are accomplished in real-world settings. Cognitive ethnography is a particularly apt method for studying instruction in both formal and informal settings. This paper discusses the practicalities of doing cognitive ethnographic research, including such issues as deciding what to record, selecting data for analysis, re-representing data, and analyzing data. Illustrative examples are provided from a recent cognitive ethnographic study of time-telling instruction.

Introduction

New technologies create new forms of inquiry. While ethnographic research has always relied on fieldnotes, the questions that could be asked and answered changed with the advent of tape recording, photography, and film. Today, new forms of digital media open new avenues of inquiry into human cultural activity. One promising new avenue is cognitive ethnography, an event-focused method for investigating how cognitive activities are accomplished in real-world settings. This chapter introduces cognitive ethnography and its application to the study of instruction, highlighting the issues involved in using the method while providing examples from a recent study of time-telling instruction.

Cognitive ethnography

Cognitive ethnography is rooted in traditional ethnography but differs from it in a fundamental way. Whereas traditional ethnography is concerned with the meanings that members of a cultural group create, cognitive ethnography is concerned with *how* members create those meanings. Traditional ethnographers might interview group members about kinship terms in an effort to delineate how they understand kinship relations and the effect this understanding has on the organization and functioning of their society. Cognitive ethnographers would be more likely to record and analyze episodes of activity in which kinship relations determine important outcomes such as control of valuable resources, seeking to understand how cultural models of kinship relations are brought to bear in processes of group decision-making. While traditional ethnography identifies the material and conceptual resources that make up group members' life worlds, cognitive ethnography examines how those resources are employed in cultural activity. Traditional ethnography gives us insight into the ways of thinking that define cultural groups, illuminating the vast range and diversity of human experience along with the commonalities that make us all human. Cognitive ethnography looks at process: at the moment-to-moment development of activity and its relation to sociocultural (often institutional) processes unfolding on different time scales. Traditional ethnography describes knowledge; cognitive ethnography describes how knowledge is constructed and used.

The contrast between cognitive ethnography and traditional ethnography should not imply a divorce of one from the other. A successful cognitive ethnography develops out of traditional ethnography, using the time-honored methods of participant observation, interviewing, artifact analysis, and the like to build the knowledge base through which specific episodes of activity can be interpreted and analyzed. How can we examine how material and conceptual resources are brought to bear in activity without knowing what material and conceptual resources are available? How can we reveal the meanings that participants construct without being familiar with the participants, their relations, and their prior history of interaction? Traditional ethnography informs cognitive ethnography; cognitive ethnography extends ethnographic research in the direction of process-level analyses.

As a method of inquiry, cognitive ethnography has several key roles to play in cognitive science (Hutchins, 2003). Its aim is to reveal how cognitive processes unfold in real-world settings—the very phenomena we are most interested in explaining. Cognitive ethnography can lend ecological validity to experimental studies by determining which questions are relevant and by developing more realistic, culturally appropriate tasks to use in laboratory investigations. Cognitive ethnography can also inform simulation studies by providing detailed descriptions of the phenomena we wish to simulate. In other words, cognitive ethnography is an important tool in the cognitive

scientist's toolkit, and when used in conjunction with other tools, it can boost our confidence in the relevance and validity of our findings.

Studying instruction

Because cognitive ethnography is a tool for studying situated activity, it is particularly apt for investigating the nature of instruction in real-world contexts, whether formal or informal. Instruction is a form of social interaction that distinguishes the human species. Through instruction, we are able to perpetuate the cultural practices (literacy, numeracy, and skilled performance in myriad domains) that maintain the cognitive sophistication of our societies across generations. Instruction may be formal, occurring in institutions specifically set up for that purpose (schools, colleges, training centers) or it may be informal, emerging during episodes of everyday discourse in many different settings (at work, during recreation, at home). Instruction may occur within activities that are highly structured or relatively unstructured; it commonly appears in structured activity where more-experienced participants guide the participation of those less-experienced. Instruction may be carefully planned (even rehearsed) or wholly spontaneous, but it is always, like other forms of discourse, improvised moment-to-moment in the unfolding of activity. How instruction develops from moment to moment, how it makes use of various material and conceptual resources, and how the micro-scale phenomena of instruction relate to cognitive, social, and cultural processes on broader timescales are exactly the sorts of questions that can be addressed by cognitive ethnographic research using digital video and analysis tools.

A recent cognitive ethnography of time-telling instruction (Williams, 2004) will help to highlight important methodological issues related to this kind of research. Here the researcher was concerned with questions in cognitive science related to the use of cognitive artifacts and the forms of cultural activity that perpetuate artifact-bound practices across generations. The study focused in particular on clock-reading and instruction in time-telling. Data for the study were collected during mathematics lessons in four classrooms (one 1st-grade, two 2nd-grade, and one 3rd-grade) in two elementary schools (an inner-city church school with a mixed student population and a mostly-white private school in an affluent community). The researcher brought to the study several forms of pertinent expertise: first, expertise in time-telling (as a competent adult member of society); second, knowledge of schools and teaching (as a licensed and experienced schoolteacher); and third, knowledge of relevant scientific theories and methods (as a member of the cognitive science research community). While these forms of expertise were essential background for the study, they were not sufficient for a successful cognitive ethnography. What was also needed was expertise about the particular community of practice in which the activities being studied took place. The researcher used traditional ethnographic methods—observation, participation, interviewing, collecting artifacts, etc., over a period of many months—to construct the knowledge base needed to analyze specific episodes of activity. Within this ethnographic research process, decisions were made about which events to capture in digital videorecordings. In this study, recordings were made only of lessons where time-telling was a focus of instruction. Because such instruction occurred somewhat rarely and sporadically throughout the school year, the researcher had to maintain close contact with the community in order to avoid missing the key events that were the target of the study.

Deciding what to record

Why not simply record everything to ensure full coverage and to amass a library of video data for future studies? Given sufficient resources, such a strategy might be appealing, but in practice, it tends to be both impractical and undesirable. Aside from the (admittedly declining) expense of such blanket recording, recording everything can be a poor strategy for several reasons. It rapidly produces unwieldy amounts of data, making it harder for the researcher to locate relevant episodes and to isolate the phenomena of interest. Just as importantly, it turns the ethnographic project into a documentary project, focusing effort on recording while diminishing the other forms of ethnographic participation needed to interpret the video data: finding a role within the community (albeit a peripheral one), forming relationships, gaining experience as a participant, interacting closely with other participants, building knowledge of the community and the relations among members, building knowledge of the activity, and so on. This concern can be ameliorated by having multiple researchers so that one is free to participate while the other records, but this also escalates the expenses of the study. If the study is concerned with phenomena that are low-frequency and unpredictable, then a broad use of recording would be warranted, but for most studies, blanket recording conveys the impression that the researcher lacks focus (i.e., does not have clearly defined research questions), and, indeed, the researcher can begin to lose focus as many different phenomena are recorded and reflected upon. Finally, and perhaps most importantly, it seems unlikely that such blanket video would really be of

much use for other studies. When recording activity for a cognitive ethnography, the researcher makes key decisions—often moment-to-moment—about where to place the camera, where to point it, what to include in the frame, which participants to focus on, when to zoom in for close-ups of details and when to pull back for wide shots that capture the broader context of activity, where to direct the microphone to capture talk or other sounds, and so on. These decisions depend directly upon the questions the researcher is trying to answer, and they produce a video record with a point of view and focus that is unlikely to capture what would be needed for a research study with different questions. In addition, a subsequent investigator would lack both experience in the community of practice and context derived from being present when and where the video was made. Lacking experience and context, the investigator would be forced to "read into" the video in order to construct a coherent account of events. Such "reading in" draws from whatever the investigator has in mind—the residua of past experiences, secondhand accounts, and depictions in the media; biases from preconceived notions; anticipations of what will be found; and so on—resulting in a default interpretation of events that is likely to be slightly or even wholly inaccurate. The problem is either not being able to interpret the video or over-interpreting the video through some largely imagined context.

This is not to say that video-recordings made for one study can never be used for another, or that it would be impossible to amass a library of video data for use in multiple studies. In fact, new innovations in technology—see, e.g., the discussion of DIVER (Pea *et al.*, this volume)—make it possible to use an omnidirectional camera to make general recordings of activity in a setting such as a classroom) and then to use "guided noticing" software to craft videoclips relevant to particular research projects, in effect postponing the videographer's moment-to-moment decisions until sometime after the event. A library of general classroom recordings could, for example, be used to investigate various questions about teacher or student behavior, although here, too, problems can arise due to the lack of context, limiting the types of questions that can be asked and answered. For most cognitive ethnographic studies, the most practical course of action is to make judicious use of recordings that focus on the phenomena of interest, walking the line between the twin dangers of failing to capture key phenomena and drowning in largely irrelevant data.

Returning to the study of time-telling instruction: because the researcher was concerned with how the teacher's talk and actions guide the mappings of concepts onto the clock face and help the students construct time readings, the decision was made to record time-telling lessons with the camera placed behind the students and focused on the teacher, in essence capturing the students' point of view (while also minimizing the distracting presence of the camera). The shots were framed to include relevant aspects of the setting, including sufficient background to provide a sense of place as well as the specific objects the teacher handled and referred to during the lesson. The camera was zoomed in just enough so that the teacher's actions and inscriptions could be seen clearly while still retaining some students (those nearest the teacher) in the shot, providing samples of student participation in the activity. Because students were seen mostly from behind, this setup also helped to preserve their anonymity. During recording, the camera direction and/or zoom were adjusted slightly to ensure that relevant items and participants were retained in the shot. Only when it seemed unavoidably necessary would the camera direction or zoom be changed more drastically, such as when a participant moved to another part of the room or made an inscription too small to be otherwise captured. Panning away from the main subject (in this case, the teacher) or zooming in on a detail (such as a student's writing) both increase the likelihood that some important action will fail to be captured in the video. Once the instructional portion of the lesson was finished and the students were set to work on their own (completing worksheets at their desks), the camera was switched off and the researcher began circulating in the room, observing students, questioning them about their work, and making fieldnotes that would later be used to warrant claims about the video data. This was part of the ethnographic component needed to inform and support analyses of the instructional episodes.

Selecting video data for analysis

For the sake of illustration, let's consider a single lesson in a 1st-grade class. The teacher had recently had the students practice dividing a circle into halves and then into fourths. In the recorded lesson, she introduced quarter-hours on the clock, preparing the students to read times as "quarter past" (the goal of the current lesson) and "quarter till" (the goal of the next lesson). The teacher had the students sit on a rug in front of her in order to receive instruction and then return to their seats to complete exercises in their workbooks. The instructional portion lasted about nine minutes and was recorded on digital video. The recorded portion of the lesson was later segmented into parts as shown in Table 1.

Table 1. Segments of instructional portion of time-telling lesson on reading "quarter past."

Activity	Focus	Duration (min:sec)
Presentation	Reviewing dividing a circle into halves and fourths on felt board	0:34
Presentation	Equating one fourth to one quarter by analogy to money	0:30
Presentation	Dividing the clock face into quarters	0:33
Presentation	Reading a time as "quarter past"	0:41
Group practice	Reading *a quarter past eight* with prompting	0:15
Group practice	Reading *a quarter past ten* and *a quarter past three*	0:15
Presentation	Counting on the clock face to read the time as "_ fifteen"	0:26
Group practice	Reading *five fifteen* and *a quarter past five*	0:17
Group practice	Reading *a quarter past seven* and *seven fifteen*	0:13
Individual practice	Reading *eight fifteen* and *a quarter past eight*	0:29
Individual practice	Reading *a quarter past two* and *two fifteen*	0:28
Presentation	Writing two fifteen as '2:15'	0:14
Individual practice	Reading *a quarter past four* and *four fifteen*; writing *4:15*	0:46
Individual practice	Reading *six fifteen* and *a quarter past six*; writing *6:15*	1:23
Individual practice	Reading *ten fifteen* and *a quarter past ten*; writing *10:15*	1:52

After segmenting the video, the next step in the process of cognitive ethnography is to select particular segments for detailed analysis. Here the guiding principle is to choose segments that exhibit the phenomena of interest, i.e., those pertinent to the research questions being investigated. Direct instruction from the teacher occurred mainly during the presentation portions of the lesson but also sporadically during group and individual practice when students needed additional prompting or made errors. Because the researcher was interested in how instruction guides mapping of conceptual elements onto the clock face, two especially relevant segments from presentation portions of the lesson were selected as the initial focus of analysis: 'dividing the clock face into quarters' and 'counting on the clock face to read the time as "__ fifteen."'

Re-representing video data

Once segments are selected, they need to be re-represented in a way that supports analysis. With video data, this re-representation is normally some form of transcription. Speech is conventionally transcribed as written text with or without annotations for pauses, emphasis, and the like. Transcribing gestures and other actions remains problematic. Gesture researchers often code gestures for such things as type, handshape, location, and motion. For the study of time-telling instruction, the researcher cared most about the relation of talk, gesture, and object-manipulation to one another, including how gestures coupled with objects the teacher was manipulating. For this reason, simply coding gestures in this way would have been insufficient. Instead, the researcher elected to use a multimodal transcript format inspired by Goodwin (see, e.g., 2003) in which talk was transcribed in conventional conversation analysis format with indications of emphasis, volume, vowel lengthening, pause length, and so on, and actions were represented via annotated still images from the video. Annotations of gestures were done in red, while annotations of object manipulations were done in blue. Each annotated image was linked to a box around the speech that co-occurred with that gesture or action. A sample transcript (much reduced and reproduced in grayscale) appears in Figure 1.

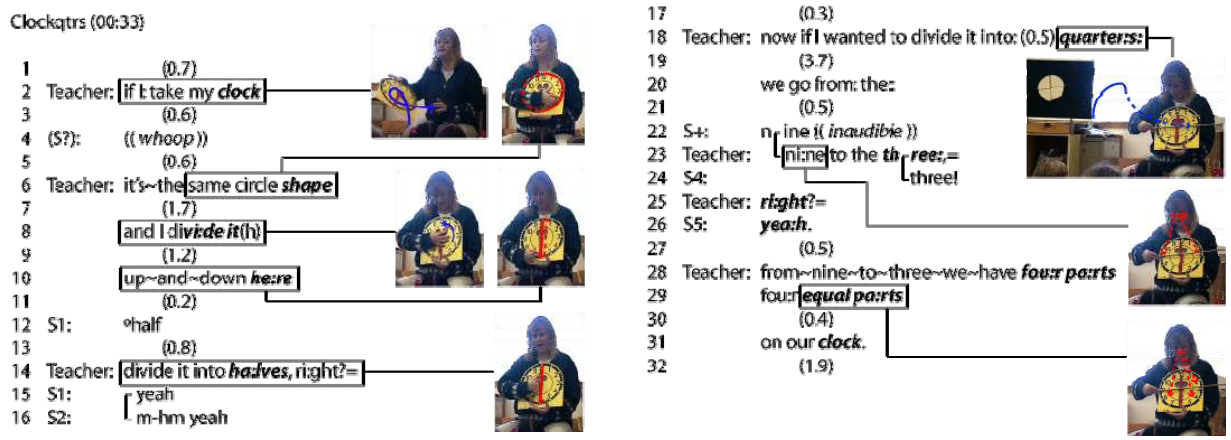

Figure 1. Transcript of the 'dividing the clock face into quarters' segment.

This multimodal transcript format has the advantages of depicting visual aspects of the discourse in visual form and clearly showing how actions were embedded in the material setting. The format has the disadvantage of leaving gestures uncoded, making it difficult to do classificatory or quantitative analyses. These multimodal transcripts are also difficult to construct, presenting issues of image choice (which image—beginning, middle, or end—best captures the gesture or action), annotation (how best to indicate form and motion, especially if these change during the gesture or action), and layout (how to fit the images and speech together on the transcript). Because of the many choices involved, multimodal transcripts typically need to be constructed by the primary researcher rather than delegated to research assistants. Fortunately, the challenging process of transcription leads to insights that inform and enrich the analyses. A more ideal form of transcript (and one that seems feasible given the state of the art in multimedia documents and web pages) would be a multimodal transcript like that shown here but with each image linked both to a videoclip showing the full gesture or action and to metadata that code the gesture by type and components, supporting other forms of gesture analysis.

Analyzing data

Looking at the transcript in Figure 1, several aspects of the instructional discourse are immediately apparent. The teacher does nearly all of the talking. She also structures both the timing and content of student talk, thus retaining control over the sequence of activity and the forms of student participation. Student talk indicates agreement or understanding or fills in pieces of information when cued by the teacher. This information may come from recall of material from previous lessons or from recognition of some state of events displayed by the teacher along with recall of an associated label. The teacher talk also follows a regular pattern: a series of single well-formed clauses or phrases ending with emphasis on a word or phrase of conceptual importance (*clock*, *shape*, *divide it*, *here*, *halves*, *quarters*, etc.) followed by a pause of a half-second or more. The effect is one of introducing one piece of information at a time and pausing while that information is processed. Here we see that although the teacher appears to control the discourse, the discourse is constrained by the students' abilities to take in and make sense of new information (i.e., by their working memory and processing capacities and present states of knowledge). The content of the discourse is shaped both by the teacher's immediate goals and by the group's shared history. A recent lesson on dividing circles, for example, becomes a resource for introducing division of the clock face. Shared history creates intersubjective awareness of states of knowledge and expectations about likely responses to particular utterances or actions; a researcher lacking domain expertise and experience as a participant-observer would be crippled in this respect.

The transcript in Figure 1 could be the basis for various analyses. Because of the structure of teacher talk (introducing one conceptual element at a time) and a research focus on how instruction guides conceptualization, the researcher elected to proceed line by line through the transcript, diagramming how each teacher utterance, gesture, or action contributed to the construction of meaning in the discourse. Such an analysis requires a conceptual framework for analyzing meaning construction. For this, the researcher drew upon the field of cognitive semantics, chiefly conceptual integration theory (Fauconnier & Turner, 2002), an account of how meaning is constructed in networks of mental spaces, and the theoretical construct of material anchors for conceptual blends (Hutchins 2005), an account of how the world is used to anchor mental spaces during reasoning or problem-solving—in essence, an

account of the functioning of various cognitive artifacts. There is not sufficient space here to review the details of these theories; for a full account, the reader is referred to the sources cited above.

Proceeding as described, the researcher began with the statement "If I take my clock" (line 2), which the teacher utters while picking up a clock face facsimile with geared hands but no mechanism for keeping time, and worked line by line to "on our clock" (line 31), the utterance that completes construction of the Clock Quarters conceptual blend. For each line of teacher talk or action, the researcher followed the conventional format of conceptual integration analyses by diagramming the mental space inputs, cross-space mappings, and resulting blended space (typically an interpretation of the clock face with particular elements as the locus of attention). The complete diagrams can be found in chapter 5 of Williams (2004); a single example will provide a flavor of the analysis. The diagram associated with line 6 of the transcript ("It's the same circle shape") is shown in Figure 2.

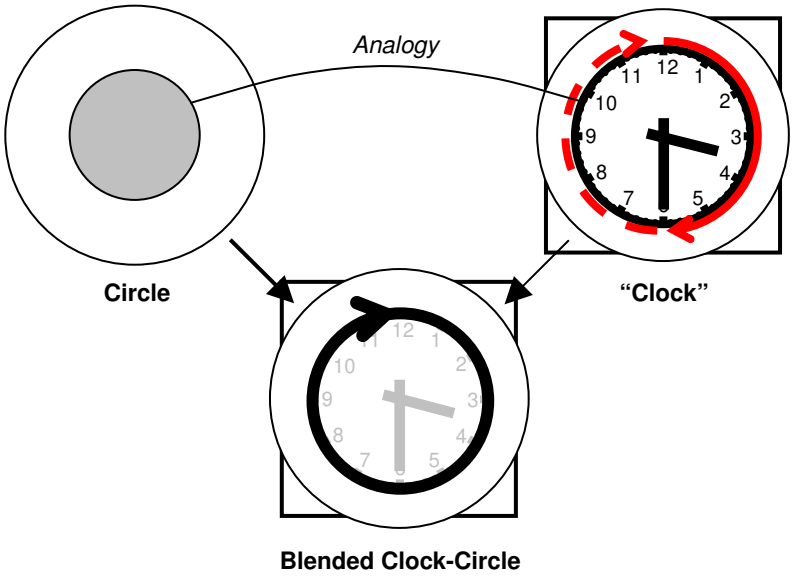

Figure 2. Conceptual integration diagram for line 6: "it's~the same circle ***shape***."

In the diagram, circles represent mental spaces and a box behind a circle shows that the mental space is anchored by structure in the world (typically by the object the students are looking at). In line 6, the word "circle" activates conceptual knowledge represented by the mental space in the upper left of the figure. "Shape" profiles a property of the clock face, represented by the mental space in the upper right anchored by the object the teacher is holding. The word "same" sets up an analogical relation (a cross-space mapping) between the conceptual circle and the clock face. The tracing gesture that coincides with "same circle shape" serves both indexical and iconic functions (see Goodwin, 2003, for discussion of this aspect of traces). Indexically, the trace highlights a portion of material structure, in this case, the clock band. Iconically, the path of the trace outlines a conceptual entity, the circle. The net effect is to superimpose the conceptual circle over its material counterpart. In the parlance of conceptual integration theory, the utterance and trace map a conceptual element onto a material anchor—in other words, they associate a conceptual entity with a material structure that will anchor it in the ensuing discourse. The resulting blended space is shown at the bottom of the figure; here the clock-circle is in profile (i.e., is the current locus of attention). In the remainder of this segment, other conceptual elements—vertical and horizontal dividing lines—are similarly mapped onto material anchors: the vertical clock hands (in a notably unclocklike configuration) and a pointing stick placed horizontally across the clock face. Again, speech profiles the conceptual element being mapped while a co-timed trace maps the conceptual element to its material anchor. In one instance (lines 20-23), the trace is executed by the eyes shifting gaze along the pointing stick from one location to another. By the end of the segment (line 31), the teacher has manipulated objects to prepare anchors for conceptual elements and used speech and co-timed gestures to map elements onto anchors, ultimately producing an anchored blended space in which the students perceive the clock face as divided into canonical clock-quarters. This blended space provides the basis for a new way to read times: as "quarter past" or "quarter till." These time readings can be constructed by reasoning within the blended space and from the blend to its associated inputs. The clock face anchors the conceptual entities

during this reasoning process, maintaining the set of conceptual relations used to generate a time reading. Indeed, this stabilizing function was a critical part of Hutchins' (2005) argument that cognitive artifacts anchor conceptual blends; what the current analysis shows is how such anchored blends get constructed in instructional discourse, making it possible for new generations to use the artifact successfully and thus keeping important cultural practices like time-telling alive.

Returning to Figure 2, we can note another important function of the gesture. Although the circle could have been traced in either direction from any starting position on the clock band (such as the location nearest the teacher's right hand), the trace started at the top of the clock face, proceeded in a clockwise direction at steady speed (continuing for a full second after the utterance), and ended when the tracing finger reached the top of the clock again. In addition to outlining a circle, the gesture also defines a path of motion that is important to time-telling: the path of the minute hand through one clock hour. In terms of cognitive semantics, the gesture imposes image-schematic structure—namely, a source-path-goal image schema—that is relevant to the current activity and that is not profiled in the accompanying speech. In other portions of the same lesson and the next day's lesson on "quarter till" (analyzed in Williams, 2004), gestures are often seen to add path information to the conceptualization.

The analysis of this excerpt would have been impoverished without knowledge gained from participant-observation, such as knowing that students had recently learned to divide a circle into halves by drawing a line down the middle and then into fourths by drawing another line across the circle from left to right—exactly the way the teacher superimposed the conceptual dividing lines on the clock face. Yet another crucial source of constraint and insight during analysis comes from juxtaposing analyses of different segments of data, looking for common patterns or contradictions. Discoveries made in one analysis often lead to re-evaluation of previous analyses. Commonalities across analyses increase both the coherence of the account and the confidence that the patterns observed in the current data are likely to generalize to instances of similar phenomena in other data. In this way, we move increasingly from descriptive to explanatory accounts.

Conclusion

As we have seen, advances in digital media have opened the way to cognitive ethnographic studies of instruction, offering new insights into the nature and function of instructional discourse. Cognitive ethnography is not for the faint-hearted, however. Because it is a tool for studying cognitive activity in real-world settings, the outcomes of a study can never be cleanly predicted. Issues of access, privacy protections, observer effects, and equipment failure loom large. Methods need to be adapted on the fly as new phenomena reveal themselves, and new forms of representation and analysis need to be developed to make sense of the data. Undertaking a cognitive ethnographic study requires an act of faith: that the phenomena of interest will appear in the data, that they will be captured in a way that supports analysis, that the findings will merit the considerable effort involved, and that we will be surprised in ways we cannot anticipate. If we believe that the phenomena we investigate are actually there in the world and that they permeate human activity, then we can be confident that once the initial shock is over—that of looking at our data and not seeing anything at all—if we persevere, new discoveries await us. That anticipation of discovery sustains us as we watch the same videos again and again, waiting for the scales to fall from our eyes and the long-invisible processes to reveal themselves.

References

Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Goodwin, C. (2003). Pointing as situated practice. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 217-242). Mahwah, NJ: Lawrence Erlbaum Associates.

Hutchins, E. (2003). Cognitive ethnography. Plenary address at the 25th annual meeting of the Cognitive Science Society.

Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics, 37*(10), 1555-1577.

Williams, R. F. (2004). *Making meaning from a clock: Material artifacts and conceptual blending in time-telling instruction*. Ph.D. dissertation, Department of Cognitive Science, University of California, San Diego.

Acknowledgments

The author is indebted to Ed Hutchins for seminal ideas related to this work. The research discussed in this chapter was supported in part by a dissertation fellowship from the Spencer Foundation.