Subverting the Comparative Research Paradigm:

The Potential for Ethnomethodology in Evaluating the Effects of Algorithm Visualization on Learning

Christopher Hundhausen
Department of Computer & Information Science
University of Oregon
Eugene, OR 97403
March, 1993

Subverting the Comparative Research Paradigm: The Potential for Ethnomethodology in Evaluating the Effects of Algorithm Visualization on Learning

1 Introduction

Predicated on the intuitive idea that one can gain insight into a computer algorithm through a mapping from its fundamental abstractions to computer graphics, algorithm visualization (AV) can be defined as the process of viewing an algorithm through a series of pictures, or through a dynamic movie, constructed both to illustrate the algorithm's dynamic behavior, and to illuminate the logic that underlies that behavior. Pioneered by Brown University's Electronic Classroom project (Van Dam 1984), and formalized by Marc Brown's seminal dissertation *Algorithm Animation* (1988), AV has gained an enthusiastic following among undergraduate computer science educators, who have come to see it as an effective and innovative method for teaching algorithms. Although computer science educators have made strong claims about AV's effectiveness (cf. Naps & Hundhausen 1991), their enthusiasm for the technique has no empirical basis. Indeed, as Stasko, Badre, & Lewis (1993) aptly point out, "the viability of algorithm [visualizations] as instructional aids remains rooted in intuition. No substantive empirical evidence has ever been presented to support these claims" (p. 1). In a culture in which the proverb "A picture is worth 10,000 words" is held so dearly, it should be no surprise that a learning technique designed to exploit the visual sense has been embraced so widely. Thus far, its powerful intuitive basis has been sufficient to justify its use.

In this paper, I address the difficult problem of evaluating AV empirically. As a starting point for my investigation, I make two narrowing assumptions: first, that there exist conditions under which AV, in some form, is not only viable as a learning aid, but also can prove to be a valuable resource in learning; and second, that it is inappropriate to advance AV systems in the absence of empirical verification or an understanding of its precise impact on learning. From this starting point, my general goal will be twofold: (1) to expose a fundamental inconcruity between the normative research paradigm, and the research goals that AV practioners have tried to meet within that paradigm; and (2) to explore the potential for an alternative research paradigm—*ethnomethodology*—both in evaluating the effects of AV on learning, and in bridging the gap that exists within the normative paradigm between methodology and goals.

In Section 2, I focus on the only study that has attempted to evaluate AV empirically (Stasko, Badre & Lewis 1993), and whose research method falls within the dominant *comparative research paradigm*[1], in which "a media-based instructional design is pitted against a non-media based design" (Williams & Brown 1990, pp. 214–215) in order to compare the new technology to traditional instructional methods. By examining both the evaluation method that the Stasko, Badre, and Lewis (hereafter SBL) study employed, and the tacit meta-assumptions that underpinned it, I shall expose the incongruities between that method, and their stated goals for the study. In so doing, I hope to answer the question "Why were the results of the SBL study disappointing?" To anticipate, an answer to that question leads us to the conclusion that the comparative research paradigm can have only limited usefulness in studying the subtle interactions that take place between an AV system and learner.

In Section 3, I present an alternative perspective on the problem—*ethnomethodology*—and I explore how that perspective might respond to the SBL study. In particular, I reevaluate the assumptions and goals of the study from the perspective of ethnomethodology, and thereby circumscribe a radically different approach to evaluating AV—one that I believe to be germane to a revised research agenda.

[1] I borrow this term from (Williams & Brown 1990).

Finally, in Section 4, I present that agenda, and I suggest some ways in which an ethnomethodological approach to the problem of evaluating AV empirically can be employed to address it.

2 The Stasko, Badre, & Lewis Study: method and meta-assumptions

In this section, I first summarize the SBL study. I then identify and scrutinize the meta-assumptions that tacitly framed the study. I conclude by applying my analysis of the SBL meta-assumptions to the question, "Why were the results of the study disappointing?"

2.1 Summary of the SBL Study

2.1.1 Background

The study focused on students' use of Tango (Transition-based ANimation GeneratiOn) (Stasko 1990), an algorithm animation[2] system Stasko developed as part of his dissertation work at Brown University in the late 1980s. Tango is driven by a mouse-based graphical interface, which allows users to interact extensively with a visualization. A visualization in Tango progresses as an animated film, with smooth transitions between dynamic program states. The particular animation that was used in this study allowed users to experiment with the *pairing heap* implementation of a *priority queue.*[3] Users could issue any of the four primary operations (insert a new node, delete node with minimum key, decrease the key of a particular node, and delete a particular node) interactively, and observe the effect of the operation on the animated display.

2.1.2 Goals

SBL set three explicit goals for their study: (1) "to examine the viability of algorithm animation as a tool for learning about a particular data structure and algorithm"; (2) "to discover the conditions under which algorithm animation could be most useful"; and (3) "to learn more about how and when to introduce algorithm animation in an educational setting" (p. 2).

2.1.3 Method

SBL gathered twenty subjectsÑall volunteers and graduate students. The subjects were evenly divided into two groups: one that would have, in addition to a couple journal articles, the benefit of Tango in learning the pairing heap implementation of a priority queue, and another that would have to rely on the journal articles alone. The control group was given 45 minutes to read about and study the pairing heap articles, whereas the animation group received 45 minutes, which it could divide between reading the articles (to a maximum of 30 minutes), and exploring the pairing heap in Tango using the interactive animation.

---

[2] Note that I will use AV ("algorithm visualization") and "algorithm animation" interchangeably in this section. Technically, however, algorithm animation is a proper subset of algorithm visualization, which encompasses any graphics-based technology—static, animated, or otherwise—intended to assist in one's understanding of the behavior of an algorithm.

[3] Note that none of the subjects had previously studied the pairing heap, which is a relatively new implementation for the priority queue., a FIFO (first in, first out) data structure in which entries may be ordered according to their priority number: entries of higher priority leave the queue before those of lower priority.

At the end of each respective learning session, subjects were given a written exam consisting of 24 questions intended to test several facets of the pairing heap algorithm. According to the authors, in order to do well on the exam, subjects needed to possess minimally (a) declarative or factual knowledge of the algorithm; (b) an understanding of the algorithm's computational complexity; and (c) procedural knowledge of the algorithm.

2.1.4 Results

Figure 1 compares the scores of the text-and-animation group to those of the text-only group. As the scores suggest, the main result of the study was that it did not reveal any significant difference between the two groups' understandings of the pairing heap algorithm. To the contrary, "while animation subjects outperformed text-only subjects, the difference was not large and was not statistically significant"[4] (ibid., p. 4). Moreover, "the [exam] performance of the animation group was not high in absolute terms" (ibid., p. 3), with an average score of only 13.6 out of a possible 24.

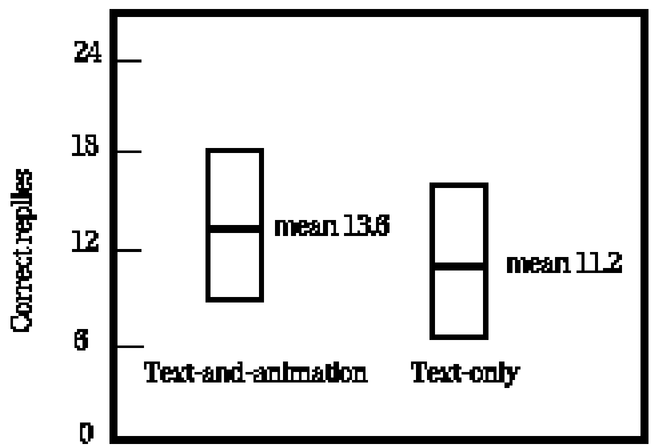

Figure 1. Comparison of test scores of text-and-animation and text-only groups [adapted from (Stasko, Badre, & Lewis 1992, p. 4)]

SBL administered a post-exam questionnaire to the animation group in order to gather some subjective, introspective data on the effects of the animation. They found that all members in the animation group believed the animation assisted in their understanding of the algorithm. Four reasons were cited. First, and most important, “the animation grabbed their interest” (ibid., p. 4). Second, they liked the interactivity of Tango; they appreciated being able to explore their own examples. Third, they found the visual effect of highlighting to be valuable in shifting their focus to the crucial portion of the animation. For example, when two nodes in the pairing heap were being compared, each node changed to a brighter color. And finally, the animation subjects appreciated the smooth transitions among animation states. As one subject put it, “I liked how the computer slowly changed the diagram, letting me see how the change was made rather than an instantaneous change” (ibid., p. 4).

On the other hand, subjects criticized the animations on four fronts. First, they longed for concomitant textual descriptions—or perhaps voice annotations—to reinforce what was going on during each stage of the animation. Second, they wanted the ability to step through the animation a frame at a time. Third, they wanted the ability to rewind the animation at any point and play it back. And finally,

---

[4] The authors performed a two sample t-test on the exam scores and found “a nonsignificant trend favoring the animation group ($t=1.111$, $df=18$, $p < 0.13$)” (Stasko, Badre, & Lewis 1992, p. 3).

several subjects noted that although they understood what was happening during an animation, they found it difficult to remember afterwards.

2.2 The meta-assumptions that underpinned the SBL study

In this section, I lay the groundwork I shall need in Section 2.3 to tackle the question " Why were the results of the SBL study disappointing?" In particular, I explicate the fundamental meta-assumptions on the basis of which I believe the SBL study was conducted. It will be illustrative to frame these assumptions in terms of three important questions that anyone studying the effects of AV on learning must address—either implicitly or explicitly. Each subsection begins with an answer to the question, and concludes with a critique of that answer.

2.2.1 What constitutes a good algorithm visualization?

Answer. In light of the authors' failure to outline (1) the attributes of the particular visual displays, (2) the interactions with the displays, and (3) the learning situations in which the interactions take place, that they believe best promote learning, we are forced to conclude that the authors are faithful to the "party line"[5] among AV practitioners. Marc Brown articulated the party line clearly in a recent article (Brown & Hershberger 1992):

> Designing an enlightening animation is a tricky psychological and perceptual challenge. What information should be presented and how? At present, creating effective dynamic visualizations of computer programs is an art, not a science. (p. 62)

To approach the question from a slightly different angle, I believe that even if AV practitioners cannot pinpoint *what* specifically constitutes a good visualization, their intuition gives them insight into *why* a good visualization assists learning. As Roschelle (1990) puts it, "computer displays can help by making experts' mental representations available in a form that students can access" (p. 2). Indeed, in accordance with Roschelle's assertion, Brown (1988) described his vision for algorithm animation thus:

> An algorithm animation environment is an exploratorium for investigating the dynamic behavior of programs, one that makes possible a fundamental improvement in the we understand and think about them. It presents multiple graphical displays of an algorithm in action, exposing properties of the program that might otherwise be difficult to understand or might even remain unnoticed. (p. 1).

Clearly, the behavior and properties that algorithm animation displays purport to expose are precisely those that experts have already come to appreciate, and that they wish to make more accessible through a mapping to graphics.

For example, consider the animation commonly employed to illustrate sorting algorithms. Each array element is represented as a stick whose height corresponds to its relative magnitude. As the array is sorted, comparisons and swaps are depicted graphically, such that two elements that are being compared are highlighted, and two elements that are being swapped smoothly switch places. Gradually, the array can be seen to fall into order, as the tops of the sticks begin to form an ascending or descending ramp. The expert can watch the animation and say, "Yes, it truly captures the essence of the algorithm," for the expert has a firm grasp of how the essential algorithm abstractions—comparisons and swaps—have been mapped to the domain of graphics. In short, the expert perceives the animation

---

5 I borrow this term from Stephen Fickas, who used it frequently in my graduate software engineering course to denote the commonly-held beliefs within a particular research community.

as a *good denotation* (Roschelle 1990) of the knowledge structures in her head. We see, then, that AV practitioners are prone to make the tacit assumption that "a good denotation of concepts will lead to appropriate knowledge structures in a learner's head" (ibid., p. 2).

Critique. Apropos of the first possible response to the question, AV practioners' relegation of AV to an art explains the heretofore lack of interest in studying the subtle interactions between student and AV system that might give rise to learning. At the same time, the attitude undermines empirical studies like SBL's, for it concedes that even if researchers were somehow able to corroborate a particular visualization's effectiveness, their findings could not be applied in any orderly fashion to the area as a whole; *just as there exists no concrete algorithm for producing good art, there exists no concrete algorithm for building good algorithm visualizations.* Hence, we have uncovered AV practioners' curiously contradictory position with respect to empirical research on AV: researchers long to establish guidelines for producing effective visualizations,[6] but their conception of AV as an *art* discourages *scientific* research into what might constitute a good visualization.

Apropos of the second possible response to the question, SBL themselves pinpoint the significant problem with assuming that a good denotation of the expert's model constitutes a good visualization: "For a student to benefit from [an] animation, the student must understand both the mapping . . . from the abstract computational algorithm domain to the animated computer graphics domain[,] . . . and the underlying algorithm upon which the mapping is based" (Stasko, Badre, & Lewis, p. 5). In other words, as faithful as that mapping may be to the expert's mental model, it is of no value to the beginner, who does not share the expert's perspective.

2.2.2 What role should an AV system play in learning?

Answer. SBL might approach this question by clarifying the role that AV should play within a larger educational context; in other words, they might answer the questions "How should AV be integrated into the curriculum," and "What is the proper mix of AV, lectures, readings, and other instructional materials?" Unfortunately, SBL do not address those important questions directly, so we are forced to infer what their answers might be from the manner in which they chose to stage their study.

Recall that subjects were divided into two groups—one of which learned the pairing heap from journal articles alone, the other of which had the benefit of the Tango animation in addition to the same journal articles. Furthermore, the latter group was required to read the articles *before* interacting with the animation. Based on these decisions, we can infer (1) that the researchers perceive AV to be a complement to, and not a replacement for, textual reading materials; and (2) that the researchers believe AV should be used to reinforce and enhance the concepts students learn about through reading; the animation sessions should take place *after* some sort of text-based introduction.

Critique. In making these ostensibly innocuous decisions about the format of the study, the researchers failed to consider a myriad of possible instructional material combinations—most notably, those involving lectures and structured laboratories, which, as researchers like Naps (1989) have argued, may be necessary concomitants of AV if it is to be effective. Insofar as the study ignored such viable

[6] cf. (Brown & Herschberger 1992), in which the authors present a list of guidelines addressing such considerations as multiple views, state cues, the effects of continuous vs. discrete transitions, multiple algorithms, selection of input data, and the use of color and sound.

combinations of learning resources, and insofar as ignoring a combination of AV and some other learning resource precluded any confirmation that AV might be viable in that combination, the study could not possibly have hoped to fulfill, in any broad sense, the first research goal that SBL set: "to examine the viability of algorithm animation as a tool for learning about a particular data structure and algorithm" (p. 2). Indeed, the study tested such a limited range of possible combinations that the researchers can only claim to have examined a tiny portion of AV's viability.

2.2.3 How will differences between learning an algorithm through AV and other media be detected?

Answer. As was mentioned in the introduction, the SBL study falls within the comparative research paradigm, which purports to measure differences in learning effectiveness by comparing and analyzing data on at least six factors that generally lend themselves to quantification: (Williams & Brown 1990, p. 214): (1) test scores, (2) heightened affective responses or better attitudes, (3) reduced learning time, (4) higher course completion rates, (5) increased retention duration, and (6) cost. In the case of the SBL study, factor (1) was directly considered, factor (2) was indirectly considered through the post-exam questionnaires,[7] and the final four factors were either dismissed as unimportant or ignored. Having relied on test scores as the primary data for their test, the authors tacitly assumed that such a measure would be sensitive enough to detect any differences in learning that each medium might promote.

Critique. It should be apparent that this assumption carries tremendous implications, the most profound of which is that the only meaningful variable in the study is the absence or presence of the interactive animation session. Indeed, the underlying assumption that SBL seem to be making is that "instructional media represent meaningful experimental variables independent of instructional content or strategy" (Williams & Brown 1990, p. 220)—what Williams and Brown have labeled the "invalid implicit assumption." In attributing the most importance to learning medium, SBL fall prey to an increasingly common tendency within the comparative research paradigm: the "tendency to treat each medium as a more or less invariant entity with fixed clusters of attributes" (ibid., p. 219). In so doing, they devalue the importance of subtle distinctions between AV and the "conventional" journal articles that may exist vis-á-vis such attributes as *content* (What information about pairing heaps did each source convey?), *language* (written language versus the visual language defined by Tango), *learner* (Do certain types of students take to AV more quickly than others?), and *learning situation* (Do certain contexts—defined by instructional goals, learning material, and participants—lend themselves to AV better than others?). Having assumed that medium was the only meaningful variable, and that the subjects' performance on a post-learning period exam was the most reliable indicator of differences in that variable, the researchers simply could not have detected subtle differences in these attributes. As a consequence, they precluded any chance of making significant progress toward meeting the second and third goals of their study: "to discover the conditions under which algorithm animations could be most useful"; and "to learn more about how and when to introduce algorithm animation in an educational setting" (p. 2).

2.3 Why were the results disappointing?

The three critiques of the meta-assumptions above all point to the same general answer to this question—namely, that the results were disappointing because the researchers' approach to the study was ill-suited to meeting their stated goals. As I see it, the most convincing evidence in support of that answer lies in SBL's own tacit acknowledgment that their statistical analyses of the two groups' test scores revealed little useful information. That tacit acknowledgment manifested itself in the Discussion section of their paper, in which the researchers concluded their analysis of the statistical results early on,

---

[7] although curiously, SBL did not give any kind of questionnaire to the text-only group.

devoting a plurality of that section to enumerating the "insights into the use of algorithm animation as an instructional aid and the conditions under which algorithm animations may be most beneficial" (p. 5) that they gleaned from the text-and-animation subjects' responses to the post-exam questionnaire. Realizing that the "hard" evidence that they had gathered—the statistical analyses of the scores—had not provided the insight they had hoped for, the researchers were forced to fall back on the qualitative questionnaire data. And working within a research paradigm that regards such data as "soft" or unscientific, SBL clearly did so at the risk of compromising the credibility of their study.

To understand why, in the midst of their disappointment, SBL found themselves having to resort to qualitative data in order to support their research goals, we need to revisit the "invalid implicit assumption" that I pointed out in Section 2.2.3. I claim that SBL's disappointment with their results is due, in large part, to the confining influence of that assumption, which served to desensitize the study to any subtle differences in learning that each source may have promoted. Recall that the researchers framed the study in such a way that (1) each learning medium (text-and-animation or text-only) was seen as an invariant entity with more or less fixed attributes; and (2) the only meaningful variable was seen as learning medium; differences in attributes such as content, language, learner, and learning situation were ignored. As a result, although the study purported to answer the questions "What are the conditions under which AV may be most useful?", and "When and how should we introduce algorithm animation in an educational setting?", the only question that the quantitative data they collected could possibly have answered is very specific: "Is there a statistically significant difference between the exam scores of two groups of students who have 45 minutes to learn an algorithm through the same journal articles, if one of those groups also has access to a Tango animation during that learning period?" It is no surprise, then, that they devoted so much attention to the qualitative data they collected; quite simply, those data actually addressed the questions they were interested in answering.

3 An ethnomethodological response to the SBL study

In the previous section, I tried to substantiate my belief that the quantitative data collected within the comparative research paradigm can do little for researchers interested in meeting the kinds of (worthy) research goals that the SBL study set. As I see it, we might obviate the problems that plague AV research conducted under that paradigm in one of two ways: either we can change our research methodology, or we can change our research goals. In this section, I address the former by outlining a radically different approach to the problem of empirically evaluating AV—one that I believe has the potential to remedy the problems that the SBL study encountered. I begin, in Section 3.1, by briefly introducing the alternative research methodology—ethnomethodology. I then clarify the foundations of the ethnomethodological approach to evaluating AV by juxtaposing them with the SBL study's methodological foundations vis-á-vis the three framing questions considered in Section 2.2. In so doing, I hope to support my belief that ethnomethodology, in remedying the methodological problems that plagued the SBL study, holds promise in meeting a revised research agenda, which I shall present in Section 4.

3.1 The perspective of ethnomethodology

Although I could not possibly do justice to the ethnomethodological perspective in this brief subsection—indeed, there exists voluminous literature on the topic that offers the same disclaimer[8]—I do need to articulate the fundamental tenets on which ethnomethodology is based. It is perhaps easiest

[8] Interestingly, a full-length book (Benson & Hughes 1983) has the same title as this subsection.

to describe those tenets by contrasting them with the tenets of conventional social science, to which, as we shall see, the SBL study tacitly subscribed.

In her captivating and highly-influential work *Plans and Situated Actions*, Suchman (1987) characterizes ethnomethodology as an "inversion" of traditional social science. To understand this inversion, we must first briefly examine the methodological premise under which social scientists have operated since the beginning of the century. Guided by the maxim "the objective reality of social facts is sociology's fundamental principle" (Suchman 1987, p. 54), social scientists have assumed that "human responses to the facts of the social world should be discoverable by the same methods as are appropriate to studies of other organisms reacting to the natural world" (ibid., p. 55). We find timely example of that assumption in the SBL study, in which the researchers assumed that the advantage of learning through AV and journal articles, as opposed to through journal articles alone, would manifest itself in the animation group's "reaction" to the medium; that reaction, they assumed, could be "measured" empirically—in much the same way that the extent of the eye's dilation can be measured when the eye is exposed to light. Drawing from early work in Ethnomethodology (Garfinkel 1967), Suchman summarizes the methodological premise of normative social science thus:

> Insofar as the alternatives of action that the culture provides are seen to be non-problematic and constraining on the individual, *their* enumeration is taken to constitute an account of situated human action. The social facts—that is to say, what actions typically come to—are used as a point of departure for retrospective theorizing about the necessary character of pathways whereby the end result is assembled. (p. 55, original emphasis)

Hence, we see that within social science's normative paradigm, the job of social scientists is to gather, substantiate, and refine the collection of social facts, along with the precise ways in which members of society are occasioned by those facts.

In contrast to the normative social science premise, ethnomethodology adopts the view that "our everyday social practices render the world publicly available and mutually intelligible" (ibid., p. 57). In other words, "objectivity is a product of systematic practices, or members' methods for rendering our unique experience and relative circumstances mutually intelligible" (ibid., p. 57). We see, then, the precise manner in which ethnomethodology inverts the premise of normative social science: whereas social science takes objective reality as the *basis* for our interaction, ethnomethodology treats objective reality—or mutual intelligibility—as the fundamental *result* of that interaction. In the following sections, we shall see how this radically different methodological perspective—truly a *Kuhnian* paradigm shift—brings itself to bear on the problem of studying AV empirically, and points to a more promising approach.

3.2 Response and revisions to the SBL meta-assumptions from the perspective of ethnomethodology

Given the perspective outlined above, I revisit, in this section, the three questions that framed our discussion of the SBL study in Section 2. By juxtaposing SBL's responses to these questions with the responses I believe ethnomethodology advocates, I intend both to delineate the vivid contrasts between the comparative research and ethnomethodology, and to concretize the ethnomethodological approach as it applies to the problem of empirically evaluating AV.

3.2.1 What constitutes a good visualization?

Recall that in Section 2.2.1, we uncovered two possible answers to this question: (1) AV is an art, not a science, so the question cannot be answered; and (2) an AV is good if it represents a good denotation of the expert's conceptualization of the algorithm. The former answer seems to concede that we cannot precisely identify the attributes of a good visualization; they depend intimately, this view maintains, on the extent to which they are meaningful to the viewer. The latter answer, on the other hand, expresses AV practitioners' inexorable desire to portray an algorithm in the form in which they conceptualize it; high value is placed on visualizations that experts believe accord well with their intuition. Notice that the question "Whose intuition?" is of paramount importance here; indeed, as SBL themselves point out,[9] the assumption that a visualization that accords well with the expert's intuition will accord well with the beginner's intuition is fallacious.

The response to this question that I believe ethnomethodology has made (Roschelle 1990) keys off the distinction between, on the one hand, a good denotation of the expert's mental model—what Roschelle has labeled *epistemic fidelity*; and, on the other hand, "the utility of a display [or an AV system] to be used as a resource for managing the uncertainty of meaning in conversations, particularly with respect to the construction of shared knowledge" (p. 3)—what Roschelle has labeled *symbolic mediation*. On this view, epistemic fidelity is rejected as a strong indicator of the goodness of a display for precisely the reason that SBL identify: the expert and the learner have different perspectives. As Roschelle puts it, "Coming from their own perspective, students do not necessarily know what to do, where to look, and how to make sense of a display" (p. 27)

In contrast, in assessing the value or goodness of a particular AV display, ethnomethodology takes as its starting point that "interpreting the significance of action is essentially a collaborative achievement" (Suchman 1987, p. 69). As Roschelle found in his own studies of the Envisioning Machine—an interactive computer simulation for teaching physics (Roschelle 1986)—this statement applies equally to interpreting the significance of displays. Using Roschelle's findings as a basis, we find that we can recast Suchman's methodological recommendation into a form that can be readily applied to the construction of graphical software:

> For students of [meaningful displays], . . . the observation that [display] interpretation is inherently uncertain does have a methodological consequence: namely, it recommends that we turn our focus from explaining away uncertainty in the interpretation of [displays] to identifying the resources by which the inevitable uncertainty is managed. [Suchman 1987, p. 69 ("display" substituted for "action")]

In short, the message that ethnomethodology is sending to AV designers advocates radical change: namely, to move *away* from trying to eliminate the inevitable uncertainty of visualizations by seeking good denotation, or by catering them to some set of display guidelines (cf. Brown & Hershberger 1992); and to move *toward* identifying precisely the resources that can be provided by an AV system, and incorporating those resources into new AV systems. The goal of the ethnomethodologically-revised approach, then, is to design "mediational tools" that purport "to provide resources to facilitate conversations between newcomers and knowledgeable community members by designing microworlds, activity structures, and visible representations of theoretical objects and properties" (Roschelle 1990, p. 27). Roschelle proposes five guidelines that may assist in achieving that goal: (1) minimalism, (2) persistence, (3) selective redundancy, (4) direct manipulation for communication, (5) activity fidelity; however, much research within this alternative paradigm remains to be done in order to refine and identify the characteristics of good displays and systems.

3.2.2 What role should an AV system play in learning?

In Section 2.2.2, we inferred that SBL believe AV should be used as a complement to, and not a replacement for, conventional instructional materials; and further, that SBL view AV as a tool for reinforcing and enhancing the material introduced to students textually—in articles or textbooks. The

---

[9] See Section 2.2.1.

ethnomethodological perspective, in contrast, argues that AV should serve as a resource whose value is dictated by the unique moment-to-moment circumstances of the learning situation.

On this view, any learning material is perceived to be just another resource for managing the inevitable uncertainty that arises in trying to interpret and make sense of an algorithm. While face-to-face interaction—what we call "tutoring" in the realm of education—is considered the richest such resource because it "incorporates the broadest range of possible resources for communication" (Suchman 1987, p. 69), other resources, of which AV and textbooks are prime examples, can prove more or less useful in constructing mutual intelligibility *in situ*. Although ethnomethodology seems to stop short of identifying the precise learning circumstances under which AV or any other communicative resource might be most valuable or appropriate—indeed, as Suchman points out, "ethnomethodology suggests . . . more a reformulation of the problem of [meaningful displays], and a research program, than an accomplished theory" [p. 50 ("meaningful displays" substituted for "purposeful action")]—it does emphasize the value of having a wide spectrum of resources "ready-to-hand," i.e., available to a conversation's participants in such a way that their use is non-problematic and largely transparent (ibid., p. 53). As Roschelle (1990) suggests, the absence of these resources can have dyer consequences on learning: "Lack of mediational means can devastate the abilities of even the best teacher; without a blackboard to draw on, models to demonstrate, and hands to wave, communication and learning both fail" (p. 19).

Hence, I believe that ethnomethodology responds to this question in a far less rigid and more adaptable way than SBL did. It says only that, as a resource, AV can potentially play a valuable role in constructing shared meaning *in situ*; and further, that a resource's being ready-to-hand for students and teachers in a learning situation can increase the likelihood of its being used successfully to build shared meaning.

3.2.3 How can differences between learning an algorithm through AV and other media be detected?

As we learned in Section 2.2.3, besides being grounded on some problematic methodological assumptions, SBL's answer to this question—factor analysis—proved too insensitive to detect any differences in learning. In that section, I tried to underscore my belief that the study's insensitivity was due, in large part, to its assumptions about meaningful variables. To understand how ethnomethodology might answer this question, we need briefly to review the contrast between SBL's and ethnomethodology's methodological framework. SBL conducted their study within the methodological framework of *Durheimian* social science, within which it is held that "human responses to the facts of the social world should be discoverable by the same methods as are appropriate to studies of other organisms reacting to the natural world" (Suchman 1987, p. 55). As I pointed out in Section 3.1, ethnomethodology stands in diametric contrast to that framework; indeed, it views objective reality not as the precondition to human interaction, but as its fundamental result.

Accordingly, the research methods to which ethnomethodology give rise reject stimulus-response kinds of quantification, opting instead for qualitative observation that focuses intently on the moment-to-moment, often subtle, interactions that take place in a learning situation as participants construct shared meaning. Those subtle interactions, it is argued, rely intimately on the peculiarities of a particular context; therefore, ethnomethodologists maintain, the central problem of the research methodology should be to identify the (situated) actions, including meaningful displays, that give rise to mutual intelligibility (Suchman 1987, p. 63).

Essential to this approach is the concept of indexicality—the "situated language" that "is not only anchored in, but in large part constitutes, the situation of use" (ibid., p. 63). In view of the importance of indexicality, Garfinkel argues that analysis cannot take place in the "general case," as SBL seem to have attempted:

> It is . . . the methodic, skillful, and therefore taken-for-granted practices whereby we establish . . . [shared meaning] . . . that provide for the analyzability of actions-in-context given that not only does no concept of context-in-general exist, but every use of "context" without exception is itself essentially indexical. (ibid., p. 62)

The ethnomethodological research method—called the *documentary method*—"describes the process whereby actions are taken as evidence, or 'documents,' of underlying plans or intent, which in turn fill in the sense of the actions" (ibid., p. 64). The documentary method shies away from generalization, instead aiming to characterize "specialized domains of practical activity" (ibid., p. 64) such as learning computer algorithms.

Within the domain of interactive computer software, ethnomethodologically-minded researchers (cf. Suchman 1987; Roschelle 1990; Douglas 1992) have advocated a research approach called *constructive interaction* (Miyake 1986), in which two person, same-sex teams (Douglas 1992, p. 2) are videotaped as they interact with software to complete a given exercise—often designed to expose the team to the full functionality of the software. As teams collaborate to construct shared meaning both of the procedures by which one interacts with the software, and of the underlying concepts that are conveyed through that interaction, they often make their understanding or misunderstanding of a particular task or concept inferable, if not explicit. Detailed analyses of the tapes of these learning episodes can give software designers insight into (1) the ability of their software to act as a resource in mediating and facilitating shared understanding (cf. Roschelle 1990, pp. 21–24); and (2) the points at which breakdown of the interface occurs, and possible design solutions that can remedy those breakdowns (cf. Douglas 1992, p. 3).

In view of the initial success that researchers like Suchman, Roschelle, and Douglas have had with using constructive interaction to evaluate interactive software, I believe that the prospects for applying the methodology to the problem of evaluating AV systems are indeed encouraging. In fact, some past systems to which this technique has been successfully applied bear a remarkable resemblance to AV systems such as Tango. In particular, Roschelle's (1990) Envisioning Machine is a simulation intended to teach the laws of velocity and acceleration, and Douglas's (1992) Cardiovascular Construction Kit is a simulation intended to teach the concepts of pressure and flow; both employ interactive simulation to teach the concepts of their particular domain in much the same way that AV systems like Tango employ interactive simulation to teach the underlying abstractions of a computer algorithm. Working within the paradigm of ethnomethodology, and drawing from the findings of constructive interaction research within the domain of interactive simulation, AV researchers find themselves in an excellent position to identify the subtle differences in learning algorithms that alternative learning conditions may promote, and that the comparative research paradigm simply cannot detect.

4 Conclusion: a revised research agenda for evaluating AV empirically

In focusing on SBL's attempt to evaluate AV empirically within the comparative research paradigm, I have exposed a serious incompatibility between the research methodology they employed, and the goals they set for the study: namely, that the kinds of data in which they were interested—data that might clarify the questions "What are the conditions under which AV may be most beneficial?" and "When and how should we introduce AV in an educational setting?"—cannot be gleaned from the statistical analysis of quantitative data. In addition, we have seen that the meta-assumptions that underpinned their study—most notably, the assumption that medium constituted the only meaningful variable—contributed greatly to rendering the research method insensitive to the goals of the study. In Section 3, I attempted to address the fundamental incongruity between methodology and goals that SBL encountered by exploring ethnomethodology, an alternative research paradigm that advocates an inversion in the fundamental way in which we perceive and study objective reality. By responding, from an ethnomethodological perspective, to the three essential questions that framed the SBL approach, I outlined the foundations and research methodology of a radically different approach to the problem of evaluating AV—one that, as I argued, holds the promise of obviating the methodological shortcomings to which the SBL study fell prey.

In sketching out the ethnomethodological response to the SBL study in the previous section, we discovered that the ethnomethodological approach proceeds from a recognition of the fundamental difference between the observational bases of the natural and social sciences. And, as we saw in Section 3.2.3, that fundamental difference gives rise to a research methodology radically different from comparative factor analysis—one that relies on the documentation of human situations within the domain of interest as its principal data. As one might anticipate, the basic research goal of the methodology keys on the same fundamental difference:

> A basic research goal for studies of situated action, therefore, is to explicate the relationship between structures of action and the resources and constraints afforded by physical and social circumstances. Ethnomethodology begins from the premise that we need, and have yet to produce, an adequate base of descriptions of situated human practices. Because there is no stable observational base, the social sciences are talking sciences, and achieve in texts, not elsewhere, the observability and practical objectivity of their phenomena. (p. 179)

If we cast this research goal into the domain of AV systems research, the message seems to be that we should shift our focus away from devising guidelines for AV displays, which are insensitive to the uncertainties that inevitably arise during the course of situated action, and instead focus our efforts on identifying the situational *resources* that an AV system can provide. Those resources, ethnomethodology seems to be suggesting, can play an important role in fulfilling three pedagogical tasks: (1) managing the uncertainties about the dynamic behavior of algorithms— and about how that behavior is depicted graphically in an AV display—that will inevitably arise; (2) facilitating collaborative conversations between users of an AV system about the important concepts underlying an algorithm; and (3) (ultimately) assisting users in the construction of a common ground—mutual intelligibility—on which the algorithms under study can be understood and appreciated.

I predict that if we can fulfill the research goal of ethnomethodology within the domain of AV—that is, if we can gather adequate accounts of student practices during the course of their interaction with AV systems—we will be able to learn more about the effects of AV on learning than we can within the comparative research paradigm. More importantly, in contrast to the SBL study's quantitative data, the qualitative data we gather through methods like constructive interaction hold the promise of giving us insights into how to build better AV systems. On the ethnomethodological view, such systems will take the form of mediational tools that are successful in establishing a shared understanding between learners becoming acquainted with algorithms, and experts who already understand and appreciate them.

References

Benson, D, & Hughes, J.A. (1983). *The perspective of ethnomethodology*. New York: Longman.

Brown, M. (1988). *Algorithm animation*. Cambridge, MA: The MIT Press.

Brown, M., & Hershberger, J. (1992). Color and sound in algorithm animation. *Computer* (December), 52–63.

Douglas, S.A. (1992). Using conversational analysis to discover breakdown in the design of user interfaces. *Proceedings of CSCW Workshop on Ethnographic Methods and Design*.

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

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science 10*, 151–177.

Naps, T.L. (1989). Design of a completely general algorithm visualization system (pp. 233–241). *Proceedings of the 22nd Annual Small College Computing Symposium*, Eau Claire, WI.

Naps, T.L, & Hundhausen, C.D. (1991). The evolution of an algorithm visualization system (pp. 259–266). *Proceedings of the 24th Annual Small College Computing Symposium*, Morris, MN.

Stasko, J., Badre, A., & Lewis, C. (1993). Do algorithm animations assist learning? An empirical study and analysis. To be published in *Proceedings of the 1993 SIGCHI Meetings* , Amsterdam, The Netherlands.

Roschelle, J. (1986). The Envisioning Machine: Facilitating students' reconceptualization of motion. Xerox Palo Alto Research Center Working Paper.

Roschelle, J. (1990). Designing for conversations. Paper presented at the AAAI Symposium on Knowledge-Based Environments for Learning and Teaching , Stanford, CA.

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

Van Dam, A. (1984). The electronic classroom: Workstations for teaching. *International Journal of Man-Machine Studies 21* (4), 353–363.

Williams, C.J. & Brown, W.W. (1990). A review of the research issues in the use of computer-related technologies for instruction: what do we know? *International Journal of Instructional Media 17* (3), 213–225.