CHAPTER 2

EPISTEMIC FIDELITY THEORY

> What makes a[n algorithm] visualization work in a lecture. . .is a short cognitive distance between concept and visualization. How closely the graphics on the screen correspond to the mental model is the best measure of a visualization's usefulness in class. When the mapping is direct and obvious, the representation is transparent and students "get it" right away. They can then concentrate on the ideas being illustrated, rather than the illustration itself. (Bazik, Tamassia, Reiss, & van Dam, 1998, p. 386)

A central argument of this thesis is that the failure of AV technology to enter mainstream CS education is rooted in fundamental deficiencies in a particular theoretical framework that has guided past research. This chapter lays the groundwork for the rest of the dissertation by (a) precisely identifying the particular theoretical framework at issue, (b) linking it to past research into AV technology design, evaluation, and pedagogy, and (c) presenting the empirical and pragmatic case against it.

The body of research into AV technology design, evaluation, and pedagogy has grown significantly over the past fifteen years. Price, Baecker, and Small (1993) estimate that papers on over 100 software visualization systems have been published. At least ten controlled experiments (Byrne, Catrambone, & Stasko, 1996, §2 & 3; Kann, Lindeman, & Heller, 1997; Lawrence, 1993, ch. 4–9; Stasko, Badre, & Lewis, 1993) explicitly evaluating AV technology's effectiveness have been reported. Finally, the literature contains the experience reports of at least a dozen computer science educators who have actually used the technology in their curricula (see, e.g., Bazik, Tamassia, Reiss, & van Dam, 1998; Brown, 1988, Appendix A; Brown & Sedgewick, 1984; Cox & Roman, 1994; Eisenstadt, Price, & Dominique, 1993; Gurka & Citrin, 1996).

On the surface, the *guiding principles* of the three lines of AV research appear to have little in common. Classic software engineering methodology, which holds modularity and reuse inviolate, has been the main guiding force behind AV systems research. Classic experimental psychology, with its interest in imposing tight environmental controls in order to ensure generalizability and replicability, has heavily influenced AV systems evaluation. The traditional didactic teaching practices of the Academy have shaped the use of AV technology in the classroom.

Yet, if one delves below the surface, I suggest that one finds a unified theoretical framework underlying the bulk of this seemingly heterogeneous research. That framework, which I call *Epistemic Fidelity* Theory[4], articulates a particular view of how and why AV technology is effective. The foundation of the EF view is a set of assumptions about what knowledge is, how it is acquired, and the efficacy of graphical representations in that acquisition. Below, I introduce the Allegory of the Musica as a means of highlighting the key assumptions of the theory. Through a critical review of the literature, I then illustrate the influence of the EF view on the research into AV design, evaluation, and pedagogy. Finally, I show that past empirical research into AV effectiveness fails to substantiate EF Theory's predictions. The case against EF Theory presented here motivates the need for an alternative guiding theory, which I introduce in Chapter 3.

[4] I borrow this term from Roschelle (1990), who, in turn, borrows it from Wenger (1987); see pp. 312–314.

2.1 The Allegory of Musica

Roaming the faraway land of Musica, the Musibeest (see Figure 8) was a large land mammal endowed with the ability to produce beautiful songs. Legend has it that ancient Musican civilizations held the Musibeest's songs in such esteem that they attempted to domesticate the animals. Their hope was to train them to sing the songs on demand.

Figure 8. The Musibeest

Unfortunately, over a millenium, the domestication process gradually killed off all but a precious few Musibeests. The near extinction of the Musibeest compelled the government to protect them. The Musican government thus established the Department of the Musibeest, an exclusive government agency charged with the mission of taking into captivity, and caring for, all remaining Musibeests.

The Musicans were a singing people, and all Musicans were endowed with a unique ability to sing the songs of the Musibeest. In fact, in Musica, the ability to reproduce Musibeest songs was regarded as the highest form of knowledge. Since so few Musibeests were around, Musicans had to rely on carefully-prepared recordings to learn the songs of the Musibeest. The Musicans developed a schooling system centered around such recordings, which, because of the general inaccessibility of Musibeests, only Musibeest instructors were allowed to make. Depending on a given instructor's knowledge of the Musibeest song, the recordings she created, and subsequently used in her instruction, more or less reflected the true nature of the Musibeest songs.

To teach students the songs of the Musibeest, instructors would rely extensively on lectures based on their recordings. In addition, instructors would often give students copies of the recordings to explore on their own. Students would take these recordings home and play them on their personal stereo systems. By repetitively listening to the recordings—both in lectures, and on their own—students would gradually internalize the songs of the Musibeest, allowing them to reproduce the songs with stunning accuracy.

The capstone of the Musican schooling system was a final exam. For the exam, instructors would first record students' own renditions of the songs, and then send the recordings to the Department of the Musibeest. Members of that agency would use a special device to compare systematically each student's voice against that of an actual Musibeest. Based on the closeness of the match, each student received a grade. Those who could produce highly faithful accounts of Musibeest songs achieved honorable status in Musica, and were invited to become Musibeest instructors. The best Musibeest instructors, as indicated by the accuracy with which their students could mimic Musibeest songs, might be invited to join the Department of the Musibeest. Such an invitation was the highest honor a Musican could hope for.

2.2 Key Assumptions

What does the Allegory of the Musibeest have to do with AV effectiveness? Below I present four key assumptions highlighted by the allegory: (a) The Knowledge Representation Assumption, (b) The Knowledge Flow Assumption, (c) The Graphical Medium Effectiveness Assumption, and (d) The Epistemic Fidelity Assumption. These assumptions form the foundation of EF Theory. In preparation for my critical review of the AV technology, evaluation, and pedagogy literature, I conclude by sketching out the practical implications of these assumptions—that is, just what it means for an AV to have high epistemic fidelity.

2.2.1 The Knowledge Representation Assumption

For the Musicans, knowledge is the song of the Musibeest; it exists independently of humans, but can be instantiated by humans who sing the songs. Analogously, according to the first key assumption of EF Theory is that knowledge exists independently of humans, but can be instantiated as symbolic structures in humans' heads. This assumption has its roots in an epistemological framework called Representationalism (see, e.g., Newell, 1980), as Doerry (1995) notes:

> The central tenet of Representationalism is that we carry inside our heads symbolic models, or *representations*, of the physical world and its behavior as well as of our intentions, goals, and beliefs with respect to the world, and the actions we can perform (i.e. plans) to achieve certain goals; these symbolic models serve as the basis for all reasoning and action that we perform. (p. 24)

2.2.2 The Knowledge Flow Assumption

In the land of Musica, people gain understanding—that is, an ability to reproduce Musibeest songs—by listening to high-fidelity recordings of the Musibeest. Thus, in Musica, target knowledge is encoded by a high-fidelity recording, and decoded by a listener; it essentially flows from Musibeest to Musican by way of a Musibeest instructor's recording. Analogously, EF Theory holds that an algorithm expert's knowledge is encoded in an AV, and decoded by an AV viewer. Thus, knowledge is seen to flow from expert to AV to viewer through the "conduit" (Reddy, 1977) of the visual medium.

Moreover, the encoding and decoding process of the transmission is seen to be *deterministic*; there exists a "deterministic correspondence between symbolic forms and their significance" (Doerry, 1995; see also Wenger, 1987). Consequently, a viewer's failure to obtain the knowledge structures encoded by an AV cannot be due to a misinterpretation of the signals; rather, it can only be attributed to "an error in transfer. .caused by some flaw or inadequacy in the medium" (Doerry, 1995).

2.2.3 The Graphical Medium Effectiveness Assumption

The Musicans believe that high-fidelity audio tape is endowed with an excellent ability to capture the songs of the Musibeest. Analogously, the EF view holds that graphical representations (such as AVs) are endowed with an excellent ability to support representations that closely match an expert's mental model of an algorithm—that is, the way in which an expert conceives of the algorithm and how it works. The belief is that graphical representations peel away unnecessary detail, presenting an expert's mental model at a level of abstraction that is "just right." The resulting match, it is held, is generally closer than that which can be achieved with other media.

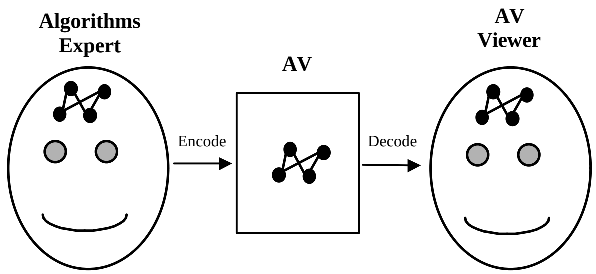

Figure 9. A Schematic Diagram of Knowledge Flow According to EF Theory

2.2.4 Overriding Assumption: Epistemic Fidelity Is Crucial

In Musican education, the more faithfully an instructor's recording captures the songs of the Musibeest, the more accurately her students will be able to reproduce Musibeest songs. Note that this conclusion logically follows from the previous three assumptions. Analogously, the overriding assumption of the EF view asserts that a close denotational match—that is, one with *high epistemic fidelity*—between an algorithm expert's knowledge structures and an AV leads to robust, efficient acquisition of those structures by the viewer, who non-problematically decodes and internalizes the target knowledge. This assumption is graphically depicted in Figure 9.

2.2.5 Practical Implications

In practical terms, what might it mean for an AV to support a high epistemic fidelity match with an expert's mental model—that is, the way in which an expert thinks about the algorithm and how it works? AV technologists emphasize that the value of a computer-based AV lies in its ability to depict *faithfully the execution of the underlying algorithm*. The assumption, as articulated by Baecker (1998), is that an AV that is "constructed automatically as a by-product of [an algorithm's] execution. .[is] guaranteed to portray this execution faithfully" (p. 369).

However, this assumption seems to overlook the fact that AVs exist as a result of human intervention. That AVs are not simply structures of logic—that they do indeed require human intervention in order to gain the putative faithfulness with which they are endowed—follows from Brown's (1988) explanation for why AVs cannot be generated automatically:

> Algorithm animation displays cannot be created automatically because they are essentially monitors of the algorithm's fundamental *operations*; an algorithm's operations cannot be deduced from an arbitrary algorithm automatically but must be denoted by *a person with knowledge of the operations performed by the algorithm*
> (p. 18, italics added)

As a member of a particular community that is interested in algorithms, the person to whom Brown refers has a sense both of what is important about algorithms, and of how to represent what is important in a way that others will understand.

We see, then, that EF Theory actually implies two distinct senses in which an AV derives epistemic fidelity:

The algorithmic sense. An AV with high epistemic fidelity maintains the intended correspondence with the dynamic behavior of the algorithm. In other words, at each point in its execution, the AV maps back to exactly that valid state in the underlying program which the AV author intends to portray. Conversely, for each state in the underlying program that is of interest to the AV author, there exists a state in the AV that depicts the interesting algorithm state as the AV author intends.

The cultural sense. An AV with high epistemic fidelity portrays the underlying algorithm in terms of the abstractions, events, and properties that expert algorithmaticians deem important and noteworthy. Furthermore, in order to represent those abstractions, events, and properties, it makes use of the notation and conventions established by expert algorithmaticans.

In short, an AV with high epistemic fidelity not only encodes the "correct" mental model of the underlying algorithm, but it does so using language that is meaningful to its viewers.

2.3 Variations on EF Theory

The version of EF Theory articulated above can be considered *strong* in the sense that it asserts that knowledge transfer occurs without regard to any other factor except epistemic fidelity. It is important to note that not all AV research subscribes to the strong version of EF Theory. Especially in the Pedagogy and Evaluation Camps, weaker versions of the theory have been explored. While the same underlying epistemology is firmly intact in these weaker versions, they hold that certain characteristics above and beyond epistemic fidelity play an important role in successful knowledge transmission. Below, I describe three different weak versions that have been considered.

2.3.1 Weak EF Theory (Learner Involvement)

Recent empirical research (Byrne, Catrambone, & Stasko, 1996; Lawrence, 1993, ch. 6 and 9) has led to the development of a weak version of EF Theory in which the AV viewer's level of attention, in addition to epistemic fidelity, plays an important role in successful knowledge transmission. Specifically, this weak version of EF Theory differs from the strong version in that it asserts that the decoding process requires the AV viewer's attention to proceed without error; the more heightened the AV viewer's attention , the more robust and efficient is the AV viewer's decoding of the AV.

Furthermore, according to weak EF Theory, heightened attention can be fostered by increasing the learner's *level of involvement*. To be more involved, a learner might do such things as construct her own data sets (Lawrence, 1993, ch. 6, 9) make explicit predictions regarding the next steps of the AV (Byrne, Catrambone, & Stasko, 1996), program an algorithm while viewing the AV (Kann, Lindeman, & Heller, 1997), or even construct her own AV (Kann, Lindeman, & Heller, 1997).

2.3.2 Weak EF Theory (Individual Differences)

A second variant of weak EF Theory ascribes significance to *individual differences*. Lawrence (1993) and Gurka and Citrin (1996) hypothesize that measurable differences in human spatial and cognitive abilities enable some individuals to decode AVs more efficiently and robustly than others. The greater the AV viewer's spatial and cognitive abilities, the more robust and efficient is the AV viewer's decoding of the target knowledge.

2.3.3 Weak EF Theory (Dual-Coding)

Based on Mayer and Anderson's (1991) *integrated dual-code hypothesis*, a third version of weak EF Theory proceeds from Paivio's (1983) assumption that "cognition consists largely of the activity of two partly interconnected but functionally independent and distinct symbolic systems" (p. 308). One encodes verbal events (words); the other encodes nonverbal events (pictures). According to Mayer and Anderson's hypothesis, AVs that encode knowledge in both verbal and non-verbal modes allow viewers to build dual *representations* in the brain, and *referential connections* between those representations. As a consequence, such AVs facilitate the transfer of target knowledge more efficiently and robustly than AVs that do not employ dual-encoding.

2.4 EF Theory's Influence on Past Research

In this section, I critically review past research into the design, evaluation, and pedagogical use of AV technology. My goal is to make the case that this body of research has pursued research agendas that have been quietly shaped by EF Theory's assumptions. In AV systems research, the strong version of the theory has been the primary guiding force; in evaluations of AV effectiveness, and in discussions of pedagogical applications, various weak versions have gained prominence. Note that whenever I refer to EF Theory below, I shall always mean the strong version of EF unless I explicitly state otherwise.

2.4.1 Influence on AV Technology Design

In Western culture, pictures in general, and animated computer graphics in particular, have widespread intuitive appeal as effective media for communication. The Graphical Medium Effectiveness Assumption (see p. 16) codifies this intuitive appeal, and was clearly a powerful motivating force behind the pioneering research into AV technology. Consider, for example, Baecker's reflections on *Sorting Out Sorting* (Baecker, 1981), the thirty minute film on sorting algorithms that is widely regarded as the seminal research of the modern AV technology era:

> The film goes beyond a step-by-step presentation of the algorithms, communicating an understanding of them as dynamic processes. We can see the programs in process, running, and we therefore see the algorithms in new and unexpected ways. We see sorting waves ripple through the data. We see data reorganize itself as if it had a life of its own. These views produce new understandings which are difficult to express in words. (Baecker, 1998, p. 377)

The advent of graphical workstations in the early 1980s, shortly after the release of *Sorting Out Sorting*, provided technologists with a medium far more interactive than film. In his seminal dissertation on graphical workstation-based AV technology, Brown (1988) is clearly as inspired as Baecker by the apparent effectiveness of computer graphics as a communication medium. Indeed, in the dissertation's opening chapter, Brown heralds the power of the graphical medium in promoting a new and improved understanding of algorithms:

> An algorithm animation environment is an "exploratorium" for investigating the dynamic behavior of programs, one that makes possible a fundamental improvement in the way 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).

Though they avoid explicit mention of knowledge representation, one might infer from quotes like the ones above that AV technologists believe knowledge to be represented by AVs. How else, one must ask, could "views produce new understandings" or "expos[e] properties of [a] program that might otherwise be difficult to understand?"

AV technology's loyalty to the remaining assumptions of EF Theory is evidenced by its direct support of specific features and techniques pioneered in Brown's (Brown, 1988; Brown & Sedgewick, 1984; Brown & Sedgewick, 1985) seminal research, and widely-embraced ever since (see, e.g., Duisberg, 1987b; Helttula, Hyrskykari, & Raiha, 1989; Roman, Cox, Wilcox, & Plun, 1992; Stasko, 1989). I detail these EF-supporting features and techniques in the following three subsections.

2.4.1.1 Support for Knowledge Flow

AV technologists' loyalty to the Knowledge Flow Assumption (see p. 16) manifests itself in at least two forms: (a) their interest in supporting increasingly high-bandwidth output media; and (b) the dyadic user model on which AV artifacts are based.

Support for increasingly high-bandwidth output. Advances in computational power over the past decade have set the stage for the development of increasingly sophisticated output technology, including color, sound, and three-dimensional (3D) graphics. AV technologists have been eager to incorporate such technology into AV systems, and to develop techniques for its use. For example, Brown and Hershberger (1992) develop techniques for incorporating color and sound into AVs. Likewise, Lieberman (1989), Roman et al. (1992), and Stasko and Wehrli (1993) present AVs that make use of the third dimension. Brown and Najork (1993) present a collection of techniques for using the third dimension in AVs.

The way in which AV technologists motivate their exploration of these new output technologies indicates a loyalty to the Knowledge Flow Assumption. In their work on three-dimensional AV, Brown and Najork (1993) sum up the motivation behind the AV Technology Camp's interest in color, sound, and 3D as follows: "The third dimension provides *an extra degree of freedom for conveying information*, much as color adds to black-and-white images, animation adds to static images, and sound adds to silent animations." (p. iv, italics added). As the quote suggests, AV technologists regard advances in output technology as opportunities to explore techniques for encoding more knowledge into AVs than was previously possible. From the strong version of EF Theory, it follows that AVs that encode more knowledge necessarily lead to the transfer of more knowledge to the AV viewer. Similarly, the *dual-coding* version of EF Theory asserts that information that is dually-encoded in both pictures and sounds, or pictures and words, leads to more robust knowledge transfer. Whether AV technologists involved in this movement were guided by strong EF or by the dual-coding version of weak EF, the influence of the Knowledge Flow Assumption on AV technolgists is clear: AV technologists are interested in exploiting color, sound, and 3D precisely because they believe those technologies will effectively increase or enhance an AV's ability to transfer knowledge.

Dyadic user model. In the opening chapter of his dissertation, Brown (1988) introduces a dyadic conceptual model for AV technology based on two conceptual actors: *client programmers*, who map algorithms to AVs; and *end-users*, who view and interact with those AVs.

According to this model, an AV system provides an *algorithm animation system* for client programmers, and an *end-user environment* for end-users. These two interfaces are not only conceptually distinct, but technologically distinct as well, as Brown (1988) notes: "The algorithm animation system is the code with which client-programmers interface, and the algorithm animation environment is the runtime environment that end-users see. It is the result of compiling the code that client-programmers implement with the algorithm animation system" (pp. 6–7).

Recall that, according to EF Theory, an algorithm expert encodes expert knowledge in an AV, and an AV viewer decodes that knowledge. Figure 10 illustrates the obvious conduits for knowledge encryption and decryption built into Brown's conceptual model. An AV artifact's programmer interface can be seen as a mechanism for encoding expert knowledge. Conversely, an AV artifact's end-user environment can be seen as a place for decoding expert knowledge.

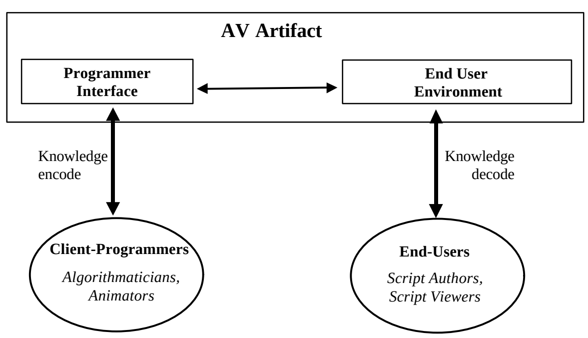

Figure 10. Knowledge Flow in Brown's (1988) User Model for AV Software

Given the clear avenues for knowledge flow provided by Brown's conceptual model, only two questions remain with respect to its adherence to the Knowledge Flow Assumption:

1. Which version of EF Theory (strong or one of the weak versions) does Brown's conceptual model support?

2. Must client-programmers necessarily be algorithm experts, and must end-users necessarily be learners? In other words, to what extent does Brown's (1988) conceptual model enforce a flow of knowledge *from* expert *to* learner?

To answer these questions, let us consider the activities that client-programmers and end-users typically take on.

1. *Client-programmers.* Client programmers either annotate algorithms to be visualized with markers "indicating interesting phenomena that should give rise to some type of display" (Brown, 1988, p. 12), or write code that maps those markers to AV views. Playing each of these roles requires a distinct form of expertise. Annotating an algorithm with event markers demands an understanding of just what is interesting about the algorithm. On the other hand, designing and implementing AV views calls for skill at computer programming, as well as detailed knowledge of the (algorithm animation system's) graphics library in terms of which AV views must be coded.

2. *End-users.* Prior to the execution of an AV, end-users may select (a) the particular AV views they wish to observe, (b) the input data on which the AV executes, and (c) the speed at which the AV is to execute. Using a *playback* interface that resembles that of a tape recorder, users may then alternatively start and pause the AV, altering the visible views and speed of execution at any point.

With respect to question (1), notice that end-users are afforded the opportunity to select input data and control AV execution. Therefore, they have the opportunity to be "involved" in the knowledge decoding process, and thus to build heightened attention. It follows that Brown's conceptual model supports the Learner Involvement version of EF Theory. With respect to question (2), notice that there exists an obvious disparity in the levels of expertise demanded of client-programmers and end-users; hence, as Brown (1988) points out, the two roles are played by different people in practice: "In an educational setting, the course instructor and teaching assistants are usually the [client-programmers]" (p. 12), while the students are most often the end-users.[5] We see, then, that Brown's conceptual model indeed supports a flow of knowledge from expert to learner, just as the Knowledge Flow Assumption prescribes.[6]

---

[5]Brown's conceptual model provides for an additional kind of end-user—the *script author.* Much as user of a word processing system can record a macro, script authors may record their AV viewing sessions—including their choice of algorithms, views, input data, and execution speed—for future replay. Hence, an exception to this rule is that instructors often play the end-user role of *script author* in order to create custom videotapes for later exploration by their students. However, as a central feature of AV end-user environments, scripting has never really caught on; since Brown's BALSA system, not one system has supported it.

[6]In apparent contradiction to the Knowledge Flow Assumption, AV technology research has exhibited a clear interest, over the past ten years, in easing the job of the client-programmer. For example, Stasko (1989; Stasko & Kraemer, 1993) demonstrates the potential for a high-level programming framework tailored specifically to algorithm animation; several AV technologists (see, e.g., Duisberg,1987a; Helttula, Hyrskykari, & Raiha, 1989; Stasko, 1989; Mukherjea & Stasko, 1994; Stasko, 1990) explore techniques for designing and implementing AVs via direct-manipulation; Najork and Brown (1995) amalgamate a high-level animation library with an interpreted language; and Citrin and Gurka (1996) advocate the use of off-the-shelf morphing software to quickly produce algorithm animations from a sequence of static images. One might construe these attempts to ease the job of client-programmers as efforts to close the knowledge gap between client-programmer and end-user, and thereby to close the knowledge flow loop set up by Brown's conceptual model. However, AV technologists involved in this movement have made it clear that their motivation lies

2.4.1.2 Support for Algorithmic Fidelity

Recall that the first form of epistemic fidelity implied by EF Theory is fidelity with an algorithm's execution. As I shall illustrate below, three features of AV technology, which support what I call *direct generation*, *input generality*, and *typeset fidelity*, were clearly designed with algorithmic fidelity in mind.

Direct generation. Noting humans' ineptitude when it comes to accurately describing dynamic algorithms, Baecker (1998) describes the obvious allure of AVs that are generated directly from executing algorithms:

> Unfortunately, a program's behavior cannot be described by a static drawing; it requires a dynamic sequence. . .It is difficult for us to enact these dynamic sequences directly. Our drawings are inaccurate. Our timing is bad. We make major mistakes, such as skipping and rearranging steps. Thus it would be useful to have animation sequences portraying the behavior of programs constructed automatically as a by-product of their execution, and therefore *guaranteed to portray this execution faithfully*. (p. 369, italics added)

The AV generation technique suggested by Baecker, which I call *direct generation*, has been nearly the only one considered by AV technologists (but see Citrin & Gurka, 1996). Indeed, so obvious seem its merits, and so tedious seem its manual alternatives, that direct generation has been a taken-for-granted constant in AV technology research ever since Baecker (1975) first advocated the technique for the production of instructional films on algorithms.

AV technologists have devoted significant time and effort to exploring alternative direct generation techniques. Early AV technology, including Brown's BALSA and Stasko's Tango, innovated so-called *annotative* techniques, in which client programmers annotate algorithm source code with interesting event procedures. When the algorithm is subsequently executed, those procedures both generate AV displays, and give rise to updates in them. The direct-manipulation techniques pioneered by Duisburg (1987a), Stasko (Mukherjea & Stasko, 1994; 1990), and Helttula (1989) are essentially variations on the annotative technique. Interesting event markers are inserted via direct manipulation, and the graphical updates to which they should give rise are specified using some combination of a graphics editor, dialog box fill-in, and direct manipulation. Finally, Roman *et al.* (1992) explore the potential for a declarative technique in which program-to-graphics mappings are specified by sets of rules. A fuller treatment of these direct generation techniques can be found in Roman and Cox's (1993) taxonomy, which uses "means of direct generation" as a key dimension for classifying the extant AV technology.

Clearly, AV technology's universal support for direct generation follows from a perceived need for high epistemic fidelity with respect to an algorithm's execution. As Baecker points out above, AVs produced by humans tend to be inaccurate; that is, they tend to have low algorithmic fidelity. Thus, if algorithmic fidelity were not such a concern for AV technologists, manually-generated AVs would suffice; direct-generation AV technology would not be necessary.

Input Generality. The generation of AVs directly from implemented computer algorithms implies that AVs, just like the algorithms on which they are based, *must operate on general input*. As Stasko (1989) points out, while it makes the client-programmer's job more difficult, such *input generality* is necessary to preserve algorithmic fidelity:

> [A]lgorithm animations. . .are especially difficult because they must proceed given any set of input data from the program. They are not one-shot animations that can be fine-tuned and discarded. They must be general enough that their animation actions convey the program's meaning for all of its possible executions. (p. 7)

---

elsewhere. Indeed, rather than perceiving a need to abandon the Knowledge Flow Assumption, researchers of this persuasion have instead consistently cited the need to make AV technology easier to use for instructors (see esp. Citrin & Gurka, 1996, pp. 2–3; Stasko,1989, p. 10).

Input generality is directly supported by Brown's (1988) notion of the *input generator*, which allows a user to specify interactively input data for an AV. Although they have not always accepted input data through a graphical-user interface like BALSA's, subsequent AV systems have consistently supported the creation of AVs that run on general input. For example, Stasko's Tango (Stasko, 1989) and Polka (Stasko & Kraemer, 1993) systems provide support for general-input AVs in the form of parameterized animation routines, which obtain their inputs directly from the underlying algorithm driving the animation.

As Stasko points out in the above quote, general-input AVs are particularly challenging to write. If AV technologists were not concerned with algorithmic fidelity, they would not be so intent on supporting general-input AVs. Instead, it seems reasonable that they would shift their focus to providing explicit mechanisms for *restricting* the range of input data. While such mechanisms would limit an AV's algorithmic fidelity, they would hold promise in simplifying the job of the AV writer—an important item on AV technologists' research agenda, as we saw above.

Typeset fidelity. Nearly all extant AV technology requires that AVs be programmed in terms *of quantitative graphics*. By quantitative graphics, I mean a graphical programming system in which (a) object position must be specified in terms of Cartesian coordinates, and (b) object attributes, as well as animation, must be specified in terms of real or integer data. The mapping of executing algorithms to quantitative graphics leads to the kinds of high-quality drawings that one would expect to find in an algorithms textbook. Because they resemble the illustrations found in typeset manuscripts, such drawings have what I call *typeset fidelity*. According to EF Theory, AVs with high typeset fidelity are attractive not only because they have the polished look of textbook illustrations (and hence *cultural fidelity*, as explained in the following subsection), but also because they are "always accurate, even the ones which would tax the best of draftsmen" (Brown, 1988, p. 5).

As was the case for direct generation and input generality, typeset fidelity does not come without formidable implementation costs for the AV designer. Unlike computers, humans do not necessarily think about graphics in terms of Cartesian coordinates; hence, laying out AVs in terms of Cartesian coordinates is a potentially time-consuming endeavor, especially when compared to, say, creating pen-and-paper sketches of AVs. That AV technologists have nonetheless insisted on facilitating the creation of precise, polished, high-typeset fidelity illustrations, rather than imprecise, scruffy, low-typeset fidelity illustrations clearly indicates an allegiance to EF Theory.

2.4.1.3 Support for Cultural Fidelity

The second form of fidelity implied by EF Theory is fidelity with the accepted conventions and nomenclature of expert algorithmaticians; one might call this fidelity with "the culture of AV." Undoubtedly inspired by illustrations in algorithms textbooks, AV technologists have been in a continuous process of developing this culture through their technology-building efforts. For example, Baecker's (1981) *Sorting Out Sorting* film firmly established the *sticks* and *dots* views of sorting algorithms; nearly two decades later, algorithmaticians refer to such views as "canonical." Building on Baecker's work, Brown and Sedgewick (1985) developed an expanded suite of AVs for sorting, searching, graph, and computational geometry algorithms; these AVs have subsequently appeared in various versions of Sedgewick's algorithms textbook (Sedgewick, 1988).

In reviewing the AV technology literature, one cannot help but notice that the views and techniques originally established by Baecker, Brown, Sedgewick, and other pioneers of AV technology have been consistently supported and extended by AV technology ever since (see, e.g., Brown & Hershberger, 1992; Duisberg, 1987b; Roman, Cox, Wilcox, & Plun, 1992; Stasko, 1989; Stasko & Kraemer, 1993). In perhaps the only attempt to document this emerging culture of AV, Brown and Hershberger (1992) concisely review the characteristics of views (state cues, history, animation transition types), and presentation techniques (multiple views, judicious input data selection, side-by-side algorithm comparison) that extant AV technology has consistently supported. That AV technologists have demonstrated a commitment to supporting the views and presentation techniques of the past suggests that the ability to support the established AV culture has become a *de facto* requirement among AV technologists.

Two lines of AV technology research, in fact, embrace this requirement explicitly. Lens (Mukherjea & Stasko, 1994) is a visualization system that allows one to create AVs rapidly through a combination of dialog box fill-in and direct manipulation. In designing Lens's visualization language, Mukherjea and Stasko (1994) studied 42 expert AVs drawn from several problem domains, including sorting, searching, and graph theory. In a similar vein, Douglas, Hundhausen, and McKeown (1995) and Chaabouni (1996) staged empirical studies involving a total of 24 algorithm experts; their interest was in designing AV languages from a user-centered (i.e., expert) perspective.

2.4.2 Influence on Effectiveness Evaluation

Nowhere is EF Theory's influence more obvious than in the arena of effectiveness evaluation. Indeed, it is through such evaluation that researchers are forced to make explicit their theory of effectiveness, which is bound up in the particulars of an empirical evaluation's design. The designs of ten controlled experiments that have explicitly tested AV's influence on learning (Byrne, Catrambone, & Stasko, 1996, §2 & 3; Kann, Lindeman, & Heller, 1997; Lawrence, 1993, ch. 4–9; Stasko, Badre, & Lewis, 1993) bear a remarkable resemblance. Below, I illustrate the way in which EF Theory has guided the design of these experiments.[7]

The Knowledge Representation Assumption asserts that individuals carry around symbolic representations of algorithms in their head. Thus, EF Theory holds that the appropriate unit of analysis for any effectiveness evaluation is individual knowledge. Accordingly, in all AV effectiveness experiments, participants have worked with AV technology individually, and their individual knowledge has been the focus of evaluation.

In particular, in line with the Knowledge Flow Assumption, effectiveness has always been operationalized in terms of the acquisition of target knowledge structures, which AV viewers are assumed to glean from learning sessions in which they are exposed to both AVs and alternative media (e.g., books, articles). Researchers have taken individual performance on a written post-test as evidence for the successful transfer of such structures, which are often classified as either *declarative* (i.e., what the program does) or *procedural* (i.e., how the algorithm works).[8]

By comparing the efficacy of various pedagogical exercises involving AV technology, all ten experiments demonstrate an implicit commitment to the Graphical Medium Effectiveness Assumption. Four of the experiments, in fact, assert the significance of the graphical medium as a causal factor in knowledge acquisition (Byrne, Catrambone, & Stasko, 1996, sec. 2 & 3; Kann, Lindeman, & Heller, 1997; Stasko, Badre, & Lewis, 1993). To see this, one need only consider their between-subjects design, as illustrated Figure 11. The underlying assumption of these experiments' design is that alternative learning *media* (see the "Expose" column in Figure 11) are more or less effective in the transmission of the target knowledge, which is subsequently quantified (see the "Measure" column Figure 11) and compared (see the "Compare" column in Figure 11).

---

[7]Gurka (1996; Gurka & Citrin, 1996) proposes a series of refinements that she claims will make experiments more sensitive to the pedagogical benefits of AV technology. However, rather than fundamentally altering the design of the experiments discussed in this section, her refinements appear simply to tweak surface features of their design.

[8]In fact, researchers have been keenly interested in ascertaining just what type of knowledge AVs successfully transfer; see, e.g., (Lawrence, 1993).

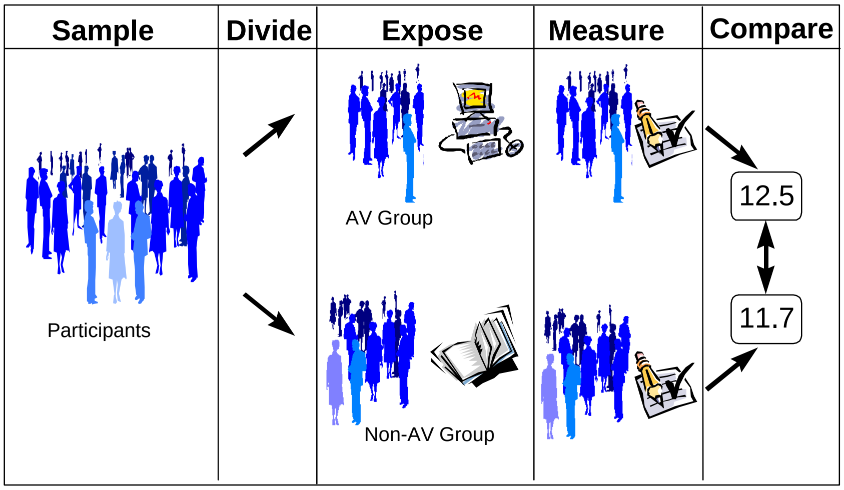

Figure 11. Prototypical Design of an AV Effectiveness Experiment

Finally, all experiments operationalize effectiveness in terms of *the robustness of individual knowledge acquisition*—the extent to which individuals acquire the target knowledge encoded by the AVs and alternative artifacts. In addition, at least one of Lawrence's (1993, ch. 5) experiments considers *knowledge acquisition time* as a dependent variable. Researchers' past interest in the robustness and speed of knowledge transfer clearly reflects a commitment to the EF Theory, which views robust knowledge transfer and efficient knowledge acquisition as the principal benefits of high-epistemic fidelity AVs.

2.4.3 Influence on Pedagogy

As I have already argued, several features of AV technology uphold EF Theory. One of those features, the dyadic user model (see Figure 10, p. 20), accommodates a set of pedagogical practices firmly rooted in EF Theory. The dyadic user encourages teachers and students to play distinct roles. Serving as client-programmers, teachers design and implement AVs. Students, on the other hand, are the end-users, viewing and interacting with those AVs. According to the Knowledge Flow Assumption, teachers *encode* the target knowledge to be learned, while students *decode* and internalize that knowledge. Its traditional advocacy of such an "encode/decode" model of learning constitutes the strongest evidence of past AV pedagogy research's adherence to EF Theory. Specifically, we find the encode/decode model of learning, as well as the EF-based assumptions underlying it, well intact in lectures, labs, and individual study—three AV-based pedagogical treatments recommended by algorithm educators.

2.4.3.1 AV in Lectures

Brown University's Electronic Classroom project pioneered the use of AV technology as a lecture aid. Brown University's Foxboro Auditorium, the original electronic classroom, contained a high-speed network of 60 high-performance graphical workstations. Using the network, instructors could broadcast AVs running on their machine to all student machines. As Brown (1988) explains, this medium of communication provided an attractive alternative to more traditional teaching media: "Rather than using chalkboards or viewgraphs to show static diagrams—often incorrectly drawn and messy at best—or asking teaching assistants to become Thespians in order to 'enact' procedure calls, searching and sorting algorithms, traveling salesmen, and so forth, dynamic simulations of the algorithms and programming concepts are presented via the workstations" (p. 165).

The structure of Brown's electronic classroom places the instructor "on stage"; each lecture can be seen as a performance in which students watch algorithm simulations both created and narrated by the instructor. As Brown University instructors note in a later experience report (Bazik, Tamassia, Reiss, & van Dam, 1998), "without interaction [among the instructor, students, and AV software], . . . a demonstration differs little, pedagogically, from a film or videotape" (pp. 3–4). It follows that, as an adjunct to such "sage-on-stage" lectures, AV technology serves as a medium for transmitting knowledge—very much in accordance with the Knowledge Flow Assumption.[9]

2.4.3.2 AV in Labs

The Denning Committee report (Denning, 1989), which recommends a core computer science curriculum, emphasizes the importance of a laboratory experience that should complement course lectures by allowing students to explore difficult concepts through experimentation. Recognizing the potential for AV technology as a tool in such a lab experience, Naps (1990) develops a pedagogy for so-called "algorithm visualization laboratories." The first principle of the pedagogy, which has been adopted by several computer science instructors since its inception, is that students should not have to implement the algorithms explored during labs. Students may have to modify an instructor-provided algorithm, but "such programming is never an end itself"; rather, "it is part of an experiment being done in the laboratory" (p. 105). A second principle of the pedagogy is that *concepts*, rather than code, should be emphasized in the lab. Implicit in this principle is the idea that AV software is ideally suited for that purpose. As Naps puts it,

> [t]he student should be able to visualize and explore the conceptual foundations of algorithms instead of the code behind them. . .[T]he amount of pure text (either code or output) seen by the student during a lab should be minimized. . .[A] lab workstation [running AV software] should be used to enhance the student's mental visualization of how the algorithm manipulates critical data structures. (p. 106)

In the AV laboratories described by Naps, the mode of interaction clearly departs from that of classroom teaching; instead of being fed with knowledge, students "explore" concepts by engaging in "experiments." However, the rationale underlying this departure from classroom teaching, as articulated by Naps in the above quote, coincides with the EF-motivated rationale behind using AV technology in lectures: Students' exposure to AVs will "enhance [their] mental visualization." The reason behind such enhancement, one can infer, is that pure textual representations of algorithms are not conceptual, while visual representations (i.e., AVs) are. Since AVs are seen to provide a closer match with target (conceptual) knowledge structures, they are to be preferred to text as the medium for target concept transfer in laboratories.

2.4.3.3 AV for individual study

A third pedagogical use of AV technology is for individual study, in which students explore AVs on their own. For example, the CD-ROM (Gloor, 1992) that accompanies the popular algorithms textbook of Cormen, Leiserson, and Rivest (1990) is ideally-suited for individual study. As Gurka and Citrin (1996) point out, providing students with AVs for individual study has both advantages and disadvantages:

> AV study aids are an improvement over unsupported studying, because correctness of the visualization is ensured, while allowing the student to produce "what-if" scenarios. . .If students are allowed complete exploratory learning, in

---

[9]According to several experience reports, the value of AV technology as a lecture aid rests in its ability to assist instructors in providing convincing on-the-spot answers to students' questions (see, e.g., Bazik, Tamassia, Reiss, & van Dam, 1998; Brown, 1988, Appendix A; Gurka & Citrin, 1996). Notice that here, too, the Knowledge Flow Assmption is at work; the only difference is that the flow of knowledge can be more responsive to students' "knowledge gaps" than it can be in non-interactive lectures.

> which they make most of the choices about examples used and their sequence, they may develop an incorrect or incomplete model due to selecting a biased problem set. (p. 183, italics added)

Once again, the rationale for AV technology's use as a study aid appears to be mired in EF Theory. One the one hand, AV is seen to provide a higher-fidelity depiction of an algorithm than other study aids. On the other hand, giving students too much freedom might lead them astray, since they may choose a "biased" problem set, leading to low epistemic fidelity AVs that cause students to acquire the wrong knowledge.

2.5 A Critique of EF Theory

The foregoing review of the AV technology literature highlights several ways in which EF Theory has quietly guided AV technologists, evaluators, and educators. To the extent that EF Theory has influenced the design, evaluation, and use of AV technology, any deficiencies in the theory have the potential to impede further research. Below, I use a review of past empirical research into AV effectiveness as a basis for critiquing the validity of the theory. As I shall illustrate, the predictions of EF Theory are not borne out by the results of past studies.

At least ten controlled experiments have explicitly evaluated the efficacy of various pedagogical treatments involving AV technology (Byrne, Catrambone, & Stasko, 1996, §2 & 3; Kann, Lindeman, & Heller, 1997; Lawrence, 1993, ch. 4–9; Stasko, Badre, & Lewis, 1993). Table 3 (next page) provides a synopsis. For each experiment cited in column 1, the pedagogical treatments that were compared appear in column 2; the measures (dependent variables) appear in column 3; and a summary of the experiment's key results appears in column 4.

Even though the experiments' treatment groups appear quite varied on the surface, further analysis indicates that all of the experiments actually manipulate just five independent variables. Table 2 (p. 28) presents those independent variables, along with the values of the variables considered by the experiments.

Notice that each of these independent variables coincides with a particular version of EF Theory. In particular, the experiments that manipulated learning medium, learning medium order, and animation representation aimed to support Strong EF Theory; the experiments that compared labeled and unlabeled representations aimed to support the Dual-coding version of Weak EF Theory; the experiment that considered spatial and verbal abilities aimed to support the Individual Differences version of Weak EF Theory; and the experiments that manipulated level of of learner involvement aimed to support the Active Learner Involvement version of Weak EF Theory. Table 3 (p. 29) reorganizes the experiments according to the versions of EF Theory they were designed to support. For each version of EF (column 1), the table lists the corresponding independent variables (column 2) and supporting experiments (column 3).

Table 1. Summary of Ten AV Effectiveness Experiments

Study	Pedagogical Treatments	Dependent Measures	Key Results
(Stasko, Badre, & Lewis, 1993)	1. Study text-only 2. Study text + view animation	1. Post-test accuracy	No significant difference
(Lawrence, 1993, ch. 4.4)	Nine treatment groups, all of which studied text and then viewed one of nine different animations formed by varying two factors: 1. Data set size (16, 27, or 41 elements) 2. Data rep. (vert. bars, horiz. bars, dots)	1. Post-test accuracy	No significant differences
(Lawrence, 1993, ch. 5)	Four treatment groups, all of which studied text and then viewed two animations formed by varying two factors: 1. Order of presentation 2. Representation style (labeled or unlabeled) (Note that spatial and verbal abilities were treated as covariates)	1. Post-test accuracy 2. Time to take post-test	1. No significant differences 2. Spatial and verbal abilities were not correlated with performance
(Lawrence, 1993, ch. 6)	1. Study text + passively view animation 2. Study text + actively view animation (by constructing own input data sets)	1. Post-test accuracy 2. Time to take post-test	Participants who actively viewed animation scored significantly higher than students who passively viewed animation
(Lawrence, 1993, ch. 7)	Four treatment groups, all of which studied text and then viewed animations formed by varying the following two factors: 1. Representation style (color vs. black-and-white 2. Representation style (algorithmic step labels vs. no labels)	1. Post-test accuracy 2. Accuracy on a transfer task	1. Participants who viewed black-and-white animations scored significantly higher on post-test 2. Participants who viewed labeled animations scored significantly higher on post-test
(Lawrence, 1993, ch. 8.2)	Four treatment groups formed by varying the following two factors: 1. Order of presentation (text first or animation first) 2. Order of presentation (selection sort first vs. Kruskal MST first)	1. Post-test accuracy	No significant differences
(Lawrence, 1993, ch. 9)	1. Lecture-only 2. Lecture + passively view animation 3. Lecture + actively view animation (by constructing own input data sets)	1. Free-response post-test accuracy 2. Multiple choice/true-false post-test accuracy	On free-response post test, participants who heard lecture and actively viewed animation significantly outperformed students who only heard lecture
(Byrne, Catrambone, & Stasko, 1996, §2)	1. Study text only 2. Study text + make predictions 3. Study text + view animation 4. Study text + view animation + make predictions	1. Post-test accuracy 2. Prediction accuracy	Participants who viewed animation and/or made predictions scored significantly higher on hard" questions than participants who did neither
(Byrne, Catrambone, & Stasko, 1996, §3)	Same as previous	Same as previous	Same as above, but difference was detected for "procedural" questions
(Kann, Lindeman, & Heller, 1997)	1. Program algorithm 2. Program algorithm + construct animation 3. Program algorithm + view animation 4. Program algorithm + view animation + construct animaton	1. Programming accuracy 2. Post-test accuracy	Participants who viewed animation scored significantly higher on post-test than participants who did

Table 2. Independent Variables Manipulated in Ten AV Effectiveness Experiments

Independent Variable	Values Considered
Learning medium	1. Text-only 2. Text-and-animation
Learning medium order	1. Text first 2. Animation first
Cognitive and Spatial Ability	1. Spatial ability (according to a paper-folding test) 2. Verbal ability (according to a vocabulary test)
Animation representation	1. Data set size (9, 25, or 41 elements) 2. Data representation (horizontal sticks, vertical sticks, dots) 3. Hue (color, black-and-white)
Representation redundancy	1. Data element labeling (labels, no labels) 2. Algorithm step labeling (labels, no labels)
Level of learner involvement	1. Passively view animation 2. Program algorithm while viewing animation 3. Actively view animation (construct own input data sets, make predictions) 4. Construct animation

Table 3. Ten AV Effectiveness Experiments vis-à-vis the Version of EF Theory For Which They Were Designed to Provide Evidence

EF Theory version	Independent Variable(s)	Supporting Experiments
Strong EF	Learning medium Learning medium order Animation representation	1. (Stasko, Badre, & Lewis, 1993) 2. (Lawrence, 1993, ch. 4) 3. (Lawrence, 1993, ch. 5) 4. (Lawrence, 1993, ch. 7) 5. (Lawrence, 1993, ch. 8)
Individual differences	Cognitive and spatial ability	1. (Lawrence, 1993, ch. 5)
Dual-Coding	Representation redundancy	1. (Lawrence, 1993, ch. 5) 2. (Lawrence, 1993, ch. 7)
Learner involvement	Level of learner involvement	1. (Lawrence, 1993, ch. 6) 2. (Lawrence, 1993, ch. 9) 3. (Byrne, Catrambone, & Stasko, 1996 295, §2) 4. (Byrne, Catrambone, & Stasko, 1996 295, §3) 5. (Kann, Lindeman, & Heller, 1997)

Figure 12 (p. 30) takes the analysis one step further by plotting the proportion of "successful" experiments vis-à-vis each EF Theory version. In other words, Figure 12 illustrates the proportion of experiments that detected statistically significant results in support of each EF Theory version. Insofar as those significant differences substantiate the causality of the independent variables manipulated in each experiment, and insofar as those independent variables express the underlying assumptions of each respective EF Theory version, the proportions plotted in Figure 12 furnish a crude measure of each EF Theory version's experimental support. [10]

As Figure 12 suggests, the evidence in support of each of the various versions of EF Theory varies considerably. Only one of the five experiments designed to support Strong EF Theory yielded a significant result; in all others, merely manipulating some aspect of the representation had no bearing on learning outcomes. Similarly, the Individual Differences version of Weak EF Theory was tested in one of Lawrence's (1993, ch. 5) early experiments, and then abandoned.

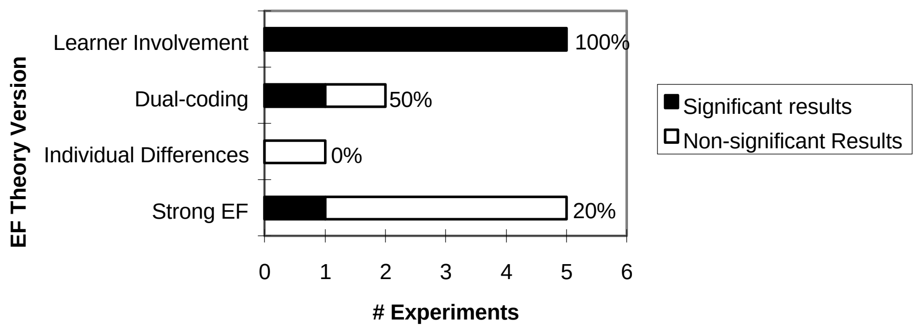

Figure 12. Summary of the Experimental Support for each Version of EF Theory

The Dual-Coding version of Weak EF Theory appears somewhat more promising than both Strong EF and the Individual Differences version of Weak EF Theory. In particular, Lawrence (1993) found that animations in which data elements are redundantly labeled do not lead to better post-test performance, whereas animations in which the algorithm's *conceptual steps* are redundantly labeled do lead to higher post-test performance. This latter result replicates an earlier experiment involving animated explanations of physical systems (Mayer & Anderson, 1991). In that experiment, animations that contained simultaneous narratives led to higher student performance on a post-test than did animations without such narratives.

Finally, the Learner Involvement version of EF Theory appears to be the most promising of the theories; indeed, it is the only one that is consistently supported by experimental evidence. In all five experiments that put this version of the theory to the test, students who were actively involved—either by programming the algorithm while viewing an animation, creating their input data and viewing an animation, or making explicit predictions while viewing an animation—performed significantly better on post-tests than those who passively watched animations (Byrne, Catrambone, & Stasko, 1996, §2 & §3; Lawrence, 1993, ch. 6), or those who were not given the opportunity to watch animations at all (Kann, Lindeman, & Heller, 1997; Lawrence, 1993, ch. 9).

Lending further credence to the Learner Involvement version of EF Theory is the historical evolution of empirical studies and experiments involving AV technology (see Figure 13). The early studies and

---

[10]Note that meta-analytic techniques like those proposed in (Hedges & Olkin, 1985) could be used to determine the *effect size* of each independent variable—a more reliable indicator of experimental support. However, such an analysis is beyond the scope of this thesis.

experiments of Badre et al. (1991), Stasko, Badre, and Lewis (1993), and Lawrence (1993, ch. 4 and 5) explored Strong EF Theory. However, a string of non-significant results led to the theory's ultimate abandonment early on in Lawrence's (1993) dissertation research. While Lawrence (1993) intermittently considered the Individual Differences (ch. 5) and Dual-Coding (ch. 5 and 7) versions of Weak EF Theory, her experiment on viewer activity (ch. 6) marked a definite turning point in the evolution of AV empirical studies, for it yielded a significant result by explicitly manipulating Learner Involvement. The viability of the Viewer Involvement version of Weak EF Theory was thus established.

From that point on, every single experiment (Byrne, Catrambone, & Stasko, 1996; Kann, Lindeman, & Heller, 1997) and empirical study (see, e.g., Kehoe & Stasko, 1996; Wilson, Katz, Ingargiola, Aiken, & Hoskin, 1995) of AV technology has included learning sessions with active learner involvement. In fact, as we shall see in the next chapter, some of the most recent AV pedagogy research (Stasko, 1997) has been interested in pushing active learner involvement even further.

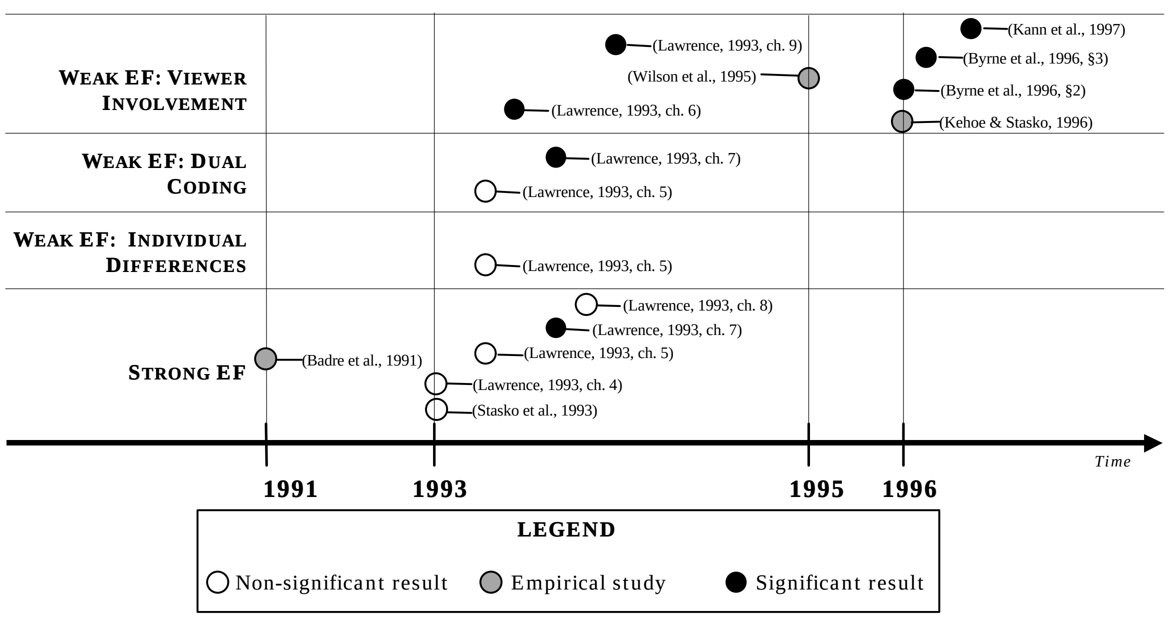

Figure 13. The Historical Evolution of AV Effectiveness Experiments and Empirical Studies

In sum, the Learner Involvement version of Weak EF Theory has garnered by far the most consistent empirical support; the historical trend in AV effectiveness studies toward active viewer involvement provides additional support for that version of the theory. Given that what AV viewers *do* has been the most significant factor in past effectiveness experiments, one must question EF Theory's assumption that an AV's epistemic fidelity matters. Indeed, the experimental results raise the possibility that an AV's epistemic fidelity matters far less than what the learner does with the AV.

2.6 Summary

In this chapter, I have described Epistemic Fidelity Theory, a particular account of what knowledge is, how it is acquired, and why AV technology is effective in facilitating knowledge acquisition. By articulating the implications of EF Theory and observing their influence on the published literature, I have shown EF Theory's stronghold on past research into AV technology, evaluation, and pedagogy.

Finally, in scrutinizing the experimental support for EF Theory, I have found definite reason to be concerned. Contrary to EF Theory's predictions, the Learner Involvement version of Weak EF Theory has garnered more consistent experimental support than any other version of EF Theory, calling into question the putative value of epistemic fidelity, and suggesting that a fundamental rethinking of our research agenda is in order. That rethinking, which I begin in the next chapter, must necessarily begin with a shift in theoretical orientation. Indeed, if we continue to allow the assumptions of EF Theory to guide our explorations of AV technology, we are likely to continue to be disappointed with the discrepancy between our intuition, which says that AV technology should be pedagogically effective, and experimental results, which indicate that what learners do matters more than what learners see.