

Cognitive Architectures for Multimedia Learning

Stephen K. Reed

*Center for Research in Mathematics & Science Education
San Diego State University*

This article provides a tutorial overview of cognitive architectures that can form a theoretical foundation for designing multimedia instruction. Cognitive architectures include a description of memory stores, memory codes, and cognitive operations. Architectures that are relevant to multimedia learning include Paivio's dual coding theory, Baddeley's working memory model, Engelkamp's multimodal theory, Sweller's cognitive load theory, Mayer's multimedia learning theory, and Nathan's ANIMATE theory. The discussion emphasizes the interplay between traditional research studies and instructional applications of this research for increasing recall, reducing interference, minimizing cognitive load, and enhancing understanding. Tentative conclusions are that (a) there is general agreement among the different architectures, which differ in focus; (b) learners' integration of multiple codes is underspecified in the models; (c) animated instruction is not required when mental simulations are sufficient; (d) actions must be meaningful to be successful; and (e) multimodal instruction is superior to targeting modality-specific individual differences.

Multimedia is increasingly providing richer environments for learning by presenting information in a wide variety of different formats. This presents a challenge for both learners and instructional designers to effectively combine this information to facilitate learning. The purpose of this article is to provide a tutorial overview of cognitive architectures that can form a theoretical foundation for multimedia learning. Multimedia in this context consists of combining words and pictures, but the different formats of words and pictures allow for many possible combinations. Words can either be written or spoken and either their phonological or semantic aspects can be emphasized. Pictures can consist of static objects, graphs, manipulated objects, or animation.

The six cognitive architectures described in this article include a description of memory stores, memory codes, and cognitive operations but do not include detailed computational models. Three theories have primarily been evaluated in traditional laboratory studies: Paivio's dual coding theory, Baddeley's working memory model, and Engelkamp's multimodal theory. The other three theories have been evaluated in instructional contexts: Sweller's cognitive load theory, Mayer's multimedia learning theory, and Nathan's ANIMATE theory.

The six theories are not rivals but focus on different aspects of multimedia learning. Table 1 compares the theories and shows those aspects that are the focus of my review. An additional dimension that is not shown in Table 1 is whether performance is measured on a reasoning or a recall task. This measure is not included because it is primarily redundant with whether the cognitive architecture emphasizes short-term memory (STM) or long-term memory (LTM). Reasoning is typically studied in STM tasks and recall in LTM tasks.

After describing the architectures I use them to present some of the advantages of having multiple memory codes. I conclude by discussing five challenging questions that require further investigation.

MULTIMODAL THEORIES

Paivio's Dual Coding Theory

Paivio's dual coding theory provided an important foundation for subsequent cognitive architectures because of its distinction between verbal and visual coding of information. The study of visual imagery had been neglected for many years following the publication of Watson's (1924) book, *Behaviorism*. Paivio (1969) reinstated visual imagery as an important topic of investigation by arguing that there are two major ways a person can elaborate on material in a learning experiment. One form of elaboration emphasizes verbal associa-

Correspondence should be addressed to Stephen K. Reed, Center for Research in Mathematics & Science Education, San Diego State University, 6475 Alvarado Road, Suite 206, San Diego, CA 92120. E-mail: sreed@sunstroke.sdsu.edu

TABLE 1
Cognitive Architectures for Multimedia Learning

<i>Theorist</i>	<i>Typical Input</i>	<i>Coding</i>	<i>Memory</i>	<i>Contribution</i>
Paivio	Words Pictures	Semantic associations Visual images	Long term	Dual coding theory
Baddeley	Words Spatial material	Phonological Visual/spatial	Short term	Working memory model
Engelkamp	Action phrases	Motor programs Semantic concepts?	Long term	Multimodal theory
Sweller	Mathematics problems Diagrams	Schema construction Schema construction	Short term	Cognitive load theory
Mayer	Science text Animation	Verbal model Pictorial model	Short term/long term	Multimedia design principles
Nathan	Word problems Animation	Problem model Situation model	Short term	Constructivist feedback

tions. A word like *freedom* may result in many associations that can distinguish it from other words. The other form of elaboration is creation of a visual image to represent a picture or a word. Paivio claimed that the concrete–abstract dimension is the most important determinant of ease in forming an image. At the concrete end of the continuum are pictures, because the picture itself can be remembered as a visual image and the person does not have to generate an image. Pictures typically result in better memory than do concrete words, which usually result in better memory than abstract words (Paivio, 1969).

If visual images and verbal associations are the two major forms of elaboration, is one more effective than the other? Paivio and his colleagues found that the imagery potential of words is a more reliable predictor of learning than the association potential of words. High-imagery words are easier to learn than low-imagery words, but high-association words are not necessarily easier to learn than low-association words (Paivio, 1969). The reason images are effective, according to Paivio (1975), is that an image provides a second kind of memory code that is independent of the verbal code. Paivio's theory is called a dual coding theory because it proposes two independent memory codes, either of which can result in recall. Having two memory codes to represent an item provides a better chance of remembering that item than having only a single code.

It should be noted that dual coding theory does not propose an integration of the verbal and visual codes because the two codes are only better than a single code if they are at least partially independent. Rather, the integration occurs for the material to be learned. Dual coding theory was initially formulated for paired-associates learning in which people had to recall a response (e.g., “letter”) to a stimulus (e.g., “corner”). Learning is particularly effective for interactive visual images (e.g., a mailbox on a street corner) that can integrate the response and the stimulus (Marschark & Hunt, 1989).

Learning foreign vocabulary words is an excellent example of an educational application of dual coding theory. The challenge in this case is to overcome the abstract nature of a

foreign word by using the mnemonic keyword method. Its effectiveness is illustrated in a study by Atkinson and Raugh (1975) on the acquisition of Russian vocabulary. The keyword method divides the study of a vocabulary word into two stages. The first stage is to associate the foreign word with an English word, the keyword, which sounds approximately like some part of the foreign word. The second stage is to form a mental image of the keyword interacting with the English translation. For example, the Russian word for *building* (*zдание*) is pronounced somewhat like *zdawn-yeh*, with the emphasis on the first syllable. Using *dawn* as the keyword, one could imagine the pink light of dawn being reflected in the windows of a building. A good keyword should (a) sound as much as possible like a part of the foreign word, (b) be different from the other keywords, and (c) easily form an interactive image with the English translation.

Students in Atkinson and Raugh's (1975) study tried to learn the English translations of 120 Russian words over a 3-day period. Those in the keyword group provided the correct translations for 72% of the Russian words, compared to 46% of the words for students in the control group. This difference is particularly impressive considering that Russian was selected as a special challenge to the keyword method because the pronunciation of most Russian words is quite different from their English pronunciation. The findings therefore illustrate the instructional power of effectively using different (phonological, semantic, visual) memory codes.

Baddeley's Working Memory Model

The working memory model proposed initially by Baddeley and Hitch (1974) also distinguishes between a verbal code and a visual code. However, the verbal code emphasizes phonological information rather than the semantic information emphasized in Paivio's dual coding theory. This is not surprising because theories of LTM such as Paivio's have traditionally emphasized semantic coding and theories of STM such as Baddeley's have traditionally emphasized phonological coding (Craik & Lockhart, 1972). In addition to the importance of

phonological coding for maintaining information in STM, phonological learning is required to learn the pronunciation of words (Baddeley, Gathercole, & Papagno, 1998).

The Baddeley and Hitch model consists of three components: (a) a phonological loop responsible for maintaining and manipulating speech-based information, (b) a visuospatial sketchpad responsible for maintaining and manipulating visual or spatial information, and (c) a central executive responsible for selecting strategies and integrating information (see Figure 1a).

The model has been applied to many different tasks to investigate how the three components are used to carry out a task, such as reproducing the location of pieces on a chess board (Baddeley, 1992). As chess players attempted to memorize the pieces on the board, they performed a secondary task that was designed to limit the use of a particular component in the model. To prevent the use of the phonological loop, the players were asked to continuously repeat a word. To prevent the use of the visuospatial sketchpad, they were asked to tap a series of keys in a predetermined pattern. To prevent the use of the central executive, they were asked to produce a string of random letters at a rate of one letter per second. Producing random letters requires people to make decisions about which letter to produce next and this requirement should restrict their ability to make decisions about how to code the chess pieces into memory.

The results of the study showed that suppressing speech had no effect on the players' ability to reproduce a chess-

board, but suppressing visual and spatial processing and requiring people to generate random letters caused a marked impairment in their ability to correctly place the pieces on the board. These findings suggest that verbal coding does not play an important role in this task, but both the visuospatial sketchpad and the central executive are needed to have good memory for the chess pieces. Other research has confirmed that simply counting the number of pieces on the board, or making decisions about moves, is affected by secondary tasks that interfere with visual and spatial processing but is unaffected by secondary tasks that prevent subvocalization (Saariluoma, 1992).

A limitation of the Baddeley and Hitch (1974) model that is particularly relevant for multimedia learning is that it did not provide a means for integrating visual and verbal codes. Both the formulations by Paivio (1969) and by Baddeley and Hitch (1974) were more useful for studying the independent contributions of verbal and visual codes than for studying the integration of two codes. The placement of the central executive between the visuospatial sketchpad and the phonological loop in Figure 1a was no accident because Baddeley and Hitch initially thought that the central executive might function as a storage system where visual and verbal codes could be integrated. However, the increasing emphasis on using the central executive to control attention left their model with no way of explaining how people can combine information from different modalities.

For this reason Baddeley (2001) proposed a revised model that contained a fourth component, the episodic buffer (Figure 1b). The episodic buffer is a storage system that can integrate memory codes from different modalities such as mentally forming a visual map from verbal directions. The purpose of this new component is to serve as a limited capacity store that can integrate information from the visuospatial sketchpad and from the phonological loop, creating a multimodal code. Another change is the inclusion of LTM to develop a greater understanding of the interaction of working memory with LTM. For example, Baddeley and Andrade (2000) found evidence for the use of the visuospatial sketchpad in working memory when participants were asked to form a novel visual image. However, when participants were asked to form an image of a familiar scene, such as a local market, LTM became more important.

One can question, however, whether there is an overemphasis on episodic information in the revised model, as revealed by the labels "episodic buffer" and "episodic LTM" in Figure 1b. Tulving and Thomson (1973) distinguished between the storage of specific episodes in episodic memory and the storage of general, factual information in semantic memory. It is not clear why Baddeley (2001) did not include semantic memory in the model. For example, a child who sees a beagle for the first time could call it a dog based on generic information about dogs stored in semantic memory. However, both the emphasis on the integration of multimodal codes and the interaction between STM and LTM will make

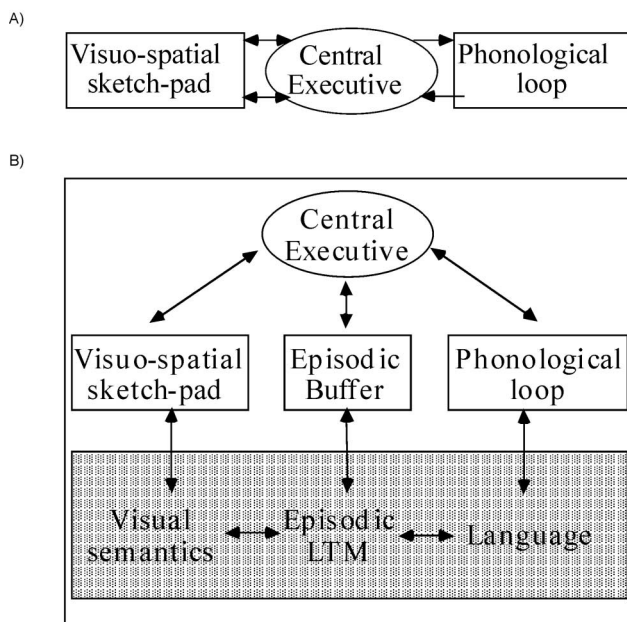


FIGURE 1 Baddeley's initial (a) and revised (b) theory of working memory. *Note.* From "Is Working Memory Working?," by A. Baddeley, 2001, *American Psychologist*, 56, p. 851–864. Copyright 2000 by the American Psychological Association. Reprinted with permission.

Baddeley's (2001) revised working memory model more relevant to multimedia learning.

Engelkamp's Multimodal Theory

The previous two architectures have focused on the interplay between words and pictures but have not incorporated actions into their design. The role that actions play in instruction has been understudied and is only recently attracting greater interest among cognitive scientists. The multimodal theory formulated by Engelkamp (1998) in his book *Memory for Actions* provides a theoretical framework for discussing some of this recent work.

Engelkamp formulated the multimodal theory to explain the empirical work that he and others had conducted over the previous decade. The typical experiment consisted of presenting participants with a list of 12 to 48 action phrases such as "nod your head" or "bend the wire" followed by the free recall of the phrases. The verbal task consisted of simply listening to the phrases and the self-performed task consisted of acting out the phrases using either imaginary or actual objects. The superiority of enactment was observed for many different conditions such as short versus long lists, pure versus mixed lists, and real versus imaginary objects (Engelkamp, 1998).

The multimodal theory shown in Figure 2 was formulated to explain the findings derived from many variations of this free recall paradigm. The theory differentiates between two modality-specific entry systems and two modality-specific output systems. The two input systems consist of the visual system (pictures, objects, events) and the verbal system that are inherent in the other architectures. The major new contribution is the relation of verbal and visual input to enactment and to the conceptual system.

One advantage of enacting phrases, as opposed to simply listening to them, is that enactment assures semantic process-

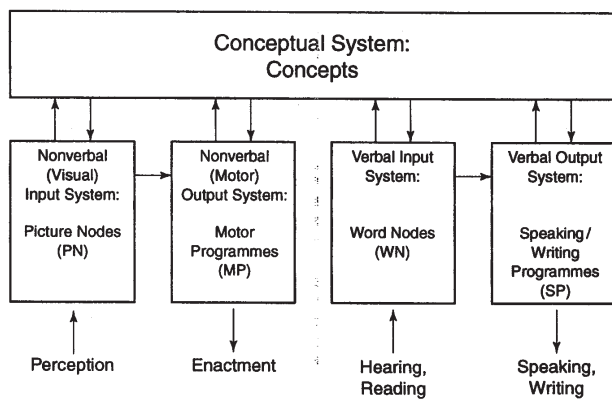


FIGURE 2 Engelkamp's multimodal theory. *Note.* From *Memory for Actions* (p. 36), by J. Engelkamp, 1998, Hove, England: Psychology Press. Copyright 1998 by Taylor & Francis. Reprinted with permission.

ing of the phrase because it is necessary to understand a command before carrying it out (Steffens, Buchner, & Wender, 2003). For example, enacting the verbal command to bend the wire provides evidence that the actor understands the command. This is illustrated in the multimodal model by the necessity of going through the conceptual system to act on a heard or read message.

Supporting evidence includes the finding that although enactment during study increases enactive clustering based on motor–movement similarities during free recall, there is also evidence of conceptual clustering based on semantic similarities (Koriat & Pearlman-Avnion, 2003). Enacting the phrase "wax the car" increased enactive clustering (e.g., "spread ointment on a wound") but there was still evidence of conceptual clustering ("pour oil into the engine"). Semantic processing of actions is also demonstrated in a paradigm in which the action is simply to press a response key by moving the index finger either toward or away from the body. The results showed an action–sentence compatibility effect in which moving one's finger in the same direction as the action phrase (e.g., toward the body for "open the drawer") resulted in faster reaction times than when the movement was incompatible with the action phrase (Glenberg & Kaschak, 2002).

INSTRUCTIONAL THEORIES

The theories proposed by Paivio, Baddeley, and Engelkamp have implications for instruction. For example, Baddeley's assumption that working memory has a limited capacity is important for multimedia learning and other forms of instruction. The instructional implications of a limited-capacity working memory have been developed in Sweller's (1988, 1994, 2003) cognitive load theory.

Sweller's Cognitive Load Theory

A potential problem in coordinating multiple representations is that the cognitive demands may overwhelm STM capacity. There are two means of overcoming this limitation through learning (Sweller, 1994). One means is through automatic processing. Automatic processing requires less memory space, freeing capacity for use elsewhere. The second means is through schema acquisition. Schemas are organized knowledge structures that increase the amount of information that can be held in working memory by chunking elements. In contrast to simple paired-associates learning, schemas are needed for more complex kinds of knowledge such as using several equations to solve a physics or a geometry problem (Sweller, 1988).

Sweller (1994) used the interactivity of schema elements to make the distinction between extraneous and intrinsic cognitive load. Intrinsic load occurs when there is high interactivity among elements in the material, so the instructional designer is unable to reduce the complexity. Extrane-

ous load occurs when the instructional designer fails to present the instructional material in a less demanding manner.

Extraneous cognitive load is important for multimedia design because the cognitive effort required to mentally integrate disparate sources of information may be reduced by physically integrating the information. For example, when studying a geometrical proof, students often need to combine information presented in both a diagram and a text. Because it requires mental effort to combine information presented in the two representations, cognitive load can be reduced by designing worked examples that carefully relate the steps in the proof with the diagram. This can be achieved by physically integrating the text and diagram to avoid a *split-attention effect* in which learners must continually shift their attention between the two representations. Sweller and his collaborators have conducted many experiments that have shown more rapid learning of instructional materials when presented in an integrated, rather than a conventional, format (Sweller, 1994).

The split attention effect occurs when multiple sources of information refer to each other and are unintelligible in isolation. However, a diagram and text will not produce a split attention effect if the diagram is fully understandable and does not require an explanation. Providing an explanation in this case can cause a *redundancy effect* in which the additional information, rather than providing a positive or a neutral effect, interferes with learning (Sweller, 2003). If one form of instruction is adequate, providing the same information in a different form will produce an extraneous cognitive load.

Mayer's Multimedia Theory

Unlike Paivio, Baddeley, and Sweller, Mayer developed a theory specifically for multimedia learning. However, these previously discussed theories form the foundation for his own contribution, as evident by the frequent references to them in his book *Multimedia Learning* (Mayer, 2001). Mayer borrows from Paivio the proposal that information can be encoded by using either a verbal or visual code. He borrows from Baddeley the idea of a limited-capacity working memory that can be managed by an executive process. He adopts Sweller's distinc-

tion between extraneous and intrinsic cognitive load, and proposes the goal of devising ways to reduce extraneous cognitive load (see Mayer & Moreno, 2003, for a detailed discussion of reducing cognitive load in multimedia learning).

Mayer's proposed architecture is shown in Figure 3. His preferred mode of presentation is to present auditory words so they do not conflict with the visual code that is needed for pictures. The sounds are organized into a verbal model and the visual images into a pictorial model. Working memory is used to integrate the verbal model, pictorial model, and prior knowledge stored in LTM. This integration occurs frequently after receiving small amounts of information, rather than at the end of the instruction.

This architecture, and his proposed principles for multimedia design, are based on dozens of experiments by Mayer and his students. The instruction typically involved a science lesson, such as how a storm forms, or the description of some device, such as how a pump works. The instruction provided (spoken or written) text with animations. Following the instruction, students answered questions that measured both retention of facts and inferences based on those facts (transfer).

This research resulted in seven principles for the design of multimedia instruction:

1. Multimedia principle: Students learn better from words and pictures than from words alone.
2. Spatial contiguity principle: Students learn better when corresponding words and pictures are presented near, rather than far from, each other on the page or screen.
3. Temporal contiguity principle: Students learn better when corresponding words and pictures are presented simultaneously rather than successively.
4. Coherence principle: Students learn better when extraneous words, pictures, and sounds are excluded.
5. Modality principle: Students learn better from animation and narration than from animation and on-screen text.
6. Redundancy principle: Students learn better from animation and narration than from animation, narration, and on-screen text.

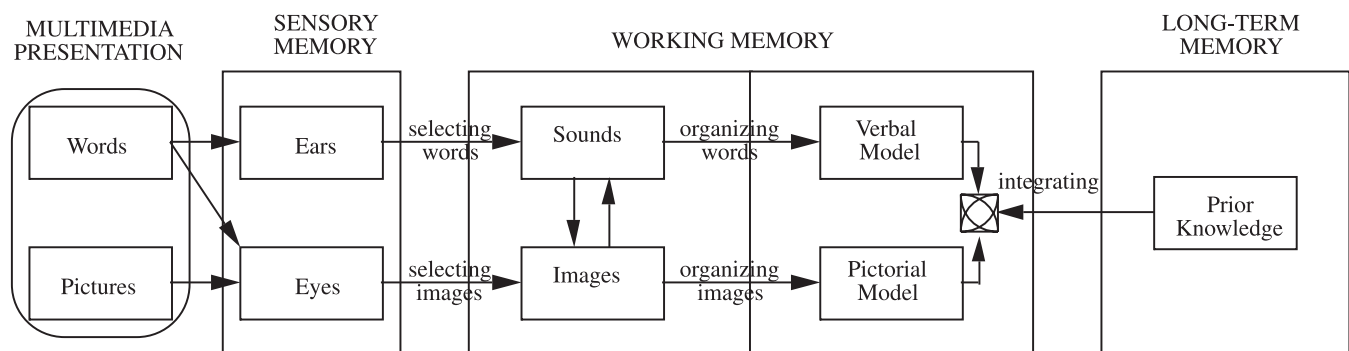


FIGURE 3 Mayer's multimedia model. Note. From *Multimedia Learning* (p. 44), by R. E. Mayer, 2001, Cambridge, England: Cambridge University Press. Copyright 2001 by Cambridge University Press. Reprinted with permission.

7. Individual differences principle: Design effects are stronger for low-knowledge learners than for high-knowledge learners and for high-spatial learners than for low-spatial learners.

Many of the principles are consistent with Mayer's goal of reducing extraneous cognitive load, such as excluding extraneous and redundant information. Presenting words and pictures near each other and in close temporal contiguity also reduces extraneous load because it increases the opportunity to have both verbal and pictorial models simultaneously available in working memory. The advantage of narration over written text is that narration and pictures occupy separate "channels" in Figure 3. Written text, like pictures, initially occupies the visual channel and then has the additional demand that it be converted back to speech to create a verbal model.

Mayer's formulation has many strengths. The model is parsimonious and easy to understand. It incorporates important ideas from the theories proposed by Paivio, Baddeley, and Sweller, and adds many research findings that specifically investigated multimedia learning. It also has practical applications, as illustrated by the seven principles for multimedia design.

Its main weakness is the underspecification of what occurs during the important integration stage in which verbal, visual, and prior knowledge are brought together in working memory. A concern raised by Schnotz (2002) is that the parallelism of text and picture processing in the model is problematic because texts and pictures are based on different sign systems that use quite different representations (see also Tabachneck-Schijf, Leonardo, & Simon, 1997). I return to the issue of integrating different modalities in the section on challenging questions.

Nathan's ANIMATE Theory

It is informative to compare the similarities and differences of another approach (Nathan, Kintsch, & Young, 1992) to multimedia instruction that combined text and animation in a somewhat different way than did Mayer. The purpose of this instruction is to use multimedia to improve students' ability to formulate equations for algebra word problems. For example, the problem illustrated in Figure 4 requires constructing an equation to find how long it will take a helicopter and a train to meet if they travel toward each other.

Students construct an equation by selecting from a palette of components those components that are part of the equation. For example, they select components specifying how to combine rate and time to find a distance; whether the two distances should be added, subtracted, or equated; and whether one time is equal, greater than, or less than the other time. Constructing a correct equation to represent a word problem depends on coordinating a situation model constructed from the text with a problem model that expresses the mathemati-

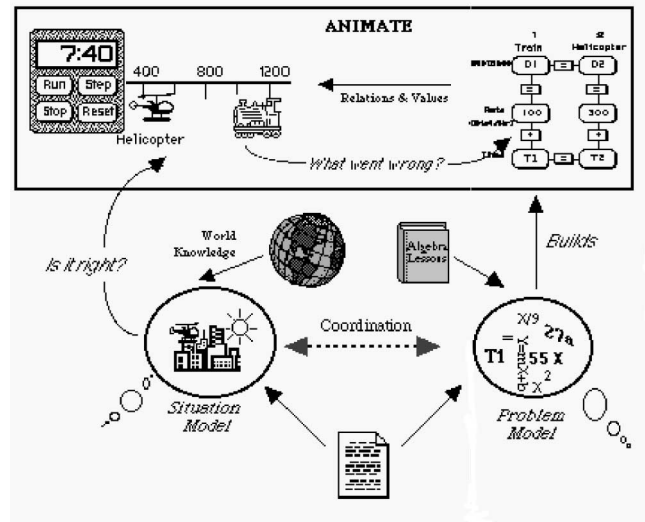


FIGURE 4 The theoretical assumptions for ANIMATE. *Note.* From "A Theory of Algebra-Word-Problem Comprehension and Its Implication for the Design of Learning Environments," by M. J. Nathan, W. Kintsch, and E. Young, 1992, *Cognition and Instruction*, 9, p. 329–389. Copyright 1992 by Lawrence Erlbaum Associates, Inc. Reprinted with permission.

cal relations among the concepts described in the situation model (see Figure 4).

A computer learning environment called ANIMATE helps to establish this correspondence by providing animation-based feedback (Nathan et al., 1992). The simulation provides visual feedback regarding whether the quantitative relations among the quantities and variables in the problem have been correctly specified. For instance, one icon may incorrectly start to move before the other, indicating an incorrect specification of the relation between the two times.

An evaluation indicated that students who used ANIMATE improved significantly more on a posttest than did students who only constructed mathematical relations without the help of animated feedback (Nathan et al., 1992). However, although many students learned, some did not (Nathan & Resnick, 1994). In these cases, additional knowledge guidance is needed. Nathan and Resnick (1994) proposed that this additional guidance can be introduced when needed to support the intended student-centered interactions.

The designers initially chose not to provide expert-guided feedback to engage the student in generating and assessing problem solutions. As in other cases of constructivist learning, the challenge is to provide the correct balance of exploration and guidance. A potential hurdle is that those students who struggle with expressing the correct mathematical relations may create relations that are impossible to simulate. For example, there are several errors in the incorrect solution shown in Figure 4. It is impossible that the helicopter and train would travel the same distance in the same amount of time if the helicopter's rate of speed is three times faster than the train's rate. It is also necessary to calculate distance by

multiplying rate and time, rather than by adding rate and time as shown in the solution. It is questionable whether an event could be simulated with so many conflicting and incorrect mathematical relations. It is therefore likely that the additional guidance will need to include verbal feedback in addition to the visual simulations.

This instruction provided by ANIMATE differs from the more traditional approach used by Mayer in which students initially combined text and pictures to form a situation model. In ANIMATE, students read a verbal description of the problem and translate it into a pictorial model of the situation, without the help of pictures. Students then evaluate the success of their mathematical constructions by determining whether the animation matches the pictorial situation model that they had constructed from reading the problem. The success of this comparison depends on students' ability to construct a correct pictorial model from the text (Kintsch, 1998), without the animation that Mayer provided to help students model the situation in his curriculum.

ADVANTAGES OF MULTIPLE CODES

The six cognitive architectures discussed in this article differ in whether they were developed to explain laboratory findings or to formulate principles for effective instruction. Paivio's (1969) and Engelkamp's (1998) formulations were developed to explain traditional research results based on paired-associates learning and free recall. Baddeley's (2001) working memory model has also been applied to many traditional laboratory tasks such as reaching conclusions from logical reasoning (Gilhooly, Logie, Wetherick, & Wynn, 1993). In contrast, the formulations developed by Sweller (1994), Mayer (2001), and Nathan et al. (1992) apply directly to instructional design. All three of these more instructionally focused models built on previous experimental work.

The interplay between traditional research and instructional applications can be seen when considering the advantages of multiple codes. One benefit is to increase recall. The advantage of having more than one kind of memory code according to Paivio's (1969) dual coding theory is that one code can serve as a backup when another code is forgotten. This same principle can be applied to all memory codes, not only the verbal and visual codes studied by Paivio. For example, the finding that self-performed actions enhance recall could be explained by the additional memory code for the motor programs that underlie the actions.

A second benefit of multiple codes is that different codes can reduce interference. Baddeley's (2001) working memory model can explain the reduction of interference if information is partitioned between the phonological loop and the visual and spatial sketchpad. An experiment that predated the Baddeley and Hitch (1974) model demonstrated that STM (working memory) could hold more information if some of the information could be stored as a verbal code and some of the

information could be stored as a visual code. People were able to recall more information if it consisted of a mix of consonants and spatial positions than if it consisted solely of consonants or solely of spatial positions (Sanders & Schroots, 1969). Mayer's (2001) modality principle—that students learn better from animation and narration than from animation and on-screen text—is another example of reducing interference.

A difference between the laboratory recall paradigms and the instructional learning paradigms is the greater need to integrate instructional information, as illustrated by the multiple sources of information in Figure 4 (Nathan et al., 1992). This need motivated the cognitive load theory proposed by Sweller (2003). Sweller's findings on physically integrating text and diagrams were verified by Mayer (2001) and form the basis for his spatial contiguity principle that students learn better when corresponding words and pictures are presented near each other. Mayer extends this idea in his temporal contiguity principle that students also learn better when corresponding narration and pictures are presented simultaneously rather than successively. The successful integration of instructional information results in a third benefit of multiple representations; they have complementary roles. By combining representations that complement each other, learners will benefit from the sum of their advantages (Ainsworth, 1999).

Another difference between the recall paradigms studied in the laboratory and instructional applications of this research is that instruction should increase both memory and understanding. Successful recall or reasoning is not always accompanied by increased understanding. As pointed out by Sweller (1988), the ability to solve problems does not necessarily imply understanding. In fact, his research has demonstrated that some problem-solving strategies require so much cognitive load that solvers have little remaining capacity to build the schematic structures that eventually lead to expert problem solving.

A fourth potential benefit of multiple codes is therefore to increase understanding (Ainsworth, 1999). Mayer (2001) distinguished between remembering and understanding by using retention tests to measure remembering and transfer tests to measure understanding. The transfer tests require that learners apply their acquired knowledge to new situations. For instance, after studying a multimedia presentation on how braking systems work, students are asked to explain how braking systems could fail. The cognitive architecture and multimedia principles developed by Mayer are based on research that demonstrates both successful retention and successful transfer. Multiple codes, however, only have the potential to increase students' understanding. Their successful use will depend on finding answers to both theoretical and instructional questions.

CHALLENGING QUESTIONS

The six cognitive architectures discussed in this article have greatly enhanced our understanding of the cognitive pro-

cesses involved in multimedia learning. Much of this research is still in its early stages, however, so it is not surprising that there are many unanswered questions. I conclude this article by discussing five questions.

Is There a Single Cognitive Architecture?

The six cognitive architectures reviewed in this article are consistent at a general level with a single cognitive architecture. For example, Sweller (2003) proposed a cognitive architecture based on the traditional model of a limited-capacity working memory combined with an unlimited capacity LTM. LTM stores schematic information that is needed for the interpretation of the modality-specific information that is held in working memory. This distinction between working memory and LTM is important in all the models, although particular models and applications may emphasize one of these memories as in Paivio's (1969, 1975) study of LTM and Baddeley's (1992, 2001) study of working memory.

At a more detailed level there are some differences in the cognitive architectures. Sweller (2003) recently argued that working memory is limited for handling new information because there is no central executive in working memory to coordinate novel information. In contrast, working memory is effective in handling previously learned material held in LTM because the previously learned material can act as a central executive. This argument obviously conflicts with the important role that the central executive has in Baddeley's (1992, 2001) model. If the central executive is defined as the component that manages information, then it is not clear why a central executive is needed to execute the well-integrated schemas stored in LTM. Rather, the central executive would be needed to manage novel information in working memory as proposed by Baddeley. Sweller's (1988) documentation of the difficulty of executing general search strategies such as means-end analysis is likely caused by the central executive's inability to keep track of the large amount of information required to select a good move.

Differences that occur among the six theories typically reflect what they attempt to model, rather than major differences in cognitive architecture. Mayer (2001) was concerned with organizing sounds into a verbal model because a spoken narrative is a primary component of his multimedia instruction. A spoken narrative is not part of the ANIMATE instruction on constructing equations, so auditory information is not part of this model (Nathan et al., 1992). Multimedia instruction is constrained by the topic being taught. A spoken narrative works well for scientific explanations but would be cumbersome when manipulating symbols and icons in the ANIMATE environment.

Mayer's and Nathan's models are examples of how instruction influences theories, but theories can also influence instruction. The instruction developed by Sweller and his collaborators is based on cognitive load theory. This instruction requires students to study a large number of worked ex-

amples to prevent cognitive overload by providing extensive guidance in solving problems. This contrasts with the constructivist approach taken in ANIMATE. According to Sweller (2003), most constructivist approaches provide too many choices to students, resulting in cognitive overload.

How Does Integration of Multiple Codes Occur?

Another commonality of the architectures is provision for the integration of codes that differ in modality. However, many theories are currently underspecified in how this integration occurs. A reasonable assumption is that integration or comparison of codes requires converting codes to a common modality. Consider an experiment by Clark and Chase (1972) that was conducted during the early years of the information-processing revolution. People had to judge whether a sentence (e.g., A is above B) correctly describes a picture. Clark and Chase argued that such comparisons required a common format, such as producing a verbal description of the picture.

More recently, Schnotz (2002) addressed this issue within the context of learning from text and visual displays. The challenge of integrating pictures and text is that they are based on different sign systems and representation principles. A text is an external descriptive representation consisting of symbols that are associated with the content they represent by convention. Visual displays are depictive representations that consist of iconic signs. However, visual displays and text can result in both internal descriptive and depictive representations by generating the internal representation that is not provided externally. This allows for the common format that may be necessary for integration and comparison of multiple codes, but does not specify which format is used.

Traditionally, the common format has been a propositional representation (*A above B*) that either is an abstract, modality-free representation or is biased toward a more verbal-based representation (Clark & Chase, 1972; Kintsch, 1998). According to Barsalou, Solomon, and Wu (1999), most representations are amodal, in which perceptual information is entered into larger symbolic structures such as frames, schemata, semantic networks, and logical expressions. In contrast, they argue for the importance of perceptual symbol systems in which the symbols are modal because they have the same structure as the perceptual states. The core assumptions of their theory are that perceptual symbols

- Represent abstract concepts directly.
- Are schematic and contain only extracted information.
- Allow the cognitive system to simulate entities and events, although these simulations are not completely unbiased and veridical.

There is now extensive behavioral and neurological evidence that mental simulations play a major role in comprehension and reasoning as documented by Barsalou (2003) and Zwan (2004). These results again raise the question of

how integration of multiple codes occurs because they underscore the importance of modality-based representations.

When Does Animation Enhance Learning?

Development of instructional software such as ANIMATE and the Animation Tutor (Reed, 2005) are based on the assumption that animation can be a major factor in enhancing learning. The recent demonstrations that mental simulations play a major role in reasoning offer both encouragement and challenges for this assumption. The encouragement is that simulation-based reasoning is a natural component of many reasoning tasks so animation-based instruction can build on this ability. The challenge is that if students can produce their own simulations, why is instructional animation necessary?

The answer to this question is that both ANIMATE and the Animation Tutor are designed to improve mathematical reasoning and problem solving by focusing on situations in which it is likely that mental simulations will require external support to be successful. ANIMATE attempts to simulate equations that students construct so they can see if their equations are correct. The Animation Tutor simulates students' estimated and calculated answers so they can judge the accuracy of their answers. Both of these simulations are beyond students' capacities to internally generate. As pointed out by Larry Barsalou (personal communication, November 9, 2004), "The issue of what happens when simulations break down is really interesting. I haven't ever thought about that, but it's clearly an important and challenging question."

Generating computer-based simulations that students cannot generate on their own is a necessary but insufficient condition for effective instructional animation. One of the limitations of animation as an instructional tool is that animation produces transient events. The research group developing cognitive load theory plans to study the implications of this limitation.

The basic hypothesis is that animation can be ineffective compared to static graphics because animation frequently turns a permanent, static graphic representation into a transient representation and the trouble with any transient representation is that WM can hold material for no more than a few seconds. Hence in this research we will be concentrating on the duration limits of WM rather than the more usual capacity limits. (J. Sweller, personal communication, December 8, 2004)

Tversky, Morrison, and Betrancourt (2002) also discussed how instructional animation may be ineffective because of design limitations. They proposed two principles to guide the construction of effective animation. The *apprehension principle* states that the content of an external representation should be accurately perceived and comprehended. Animations that allow close-ups, zooming, alternative perspectives, and control of speed are likely to facilitate perception and comprehen-

sion. Note that this recommendation supports a more interactive approach in viewing the animations. The *congruence principle* states that the structure and content of the external representation should correspond to the desired structure and content of the internal representation. For example, there should be a natural correspondence between change over time and the essential conceptual information to be conveyed.

When Do Actions Enhance Learning?

The research on recalling action phrases (Engelkamp, 1998) is relevant for education because actions will hopefully increase understanding through semantic processing. However, instructional learning typically is more organized than executing or recalling unrelated action phrases and requires the kind of schematic learning studied by Sweller, Mayer, and Nathan. An encouraging result demonstrating that action can play a helpful role in more complex instruction is the finding that gesturing reduced the cognitive demands on working memory when students were asked to explain their solutions to mathematics problems they had written on a blackboard (Wagner, Nusbaum, & Goldin-Meadow, 2004). Students were able to recall more supplementary information (either a string of letters or a visual grid) from working memory when they gestured while explaining their solution than when they did not gesture. The number of remembered items depended on the meaning of the gesture, with more items recalled when the gesture and verbal explanations conveyed the same meaning. However, the mismatch between information conveyed by gesture and by speech can provide useful diagnostic information, such as determining when students are considering different solution options as they solve problems (Garber & Goldin-Meadow, 2002).

An instructional challenge raised by Engelkamp's (1998) multimodal theory is that although translating verbal information into appropriate actions requires semantic processing, translating visual input into actions can bypass semantic processing as indicated in Figure 2. For example, a student might perform actions on manipulatives during a mathematics class without understanding their connections to the mathematics. Good manipulatives, according to Clements (1999), are those that are meaningful to the learner, provide control and flexibility, and assist the learner in making connections to cognitive and mathematical structures. His own software program called Shapes allows children to use computer tools to move, combine, duplicate, and alter shapes to make designs and solve problems.

In commenting on the mixed results of research on the effectiveness of manipulatives, Thompson (1994) proposed that it is necessary to look at the total instructional environment to understand the effectiveness of concrete materials. Although the material may be concrete, the idea the material is supposed to convey may not be obvious because of students' ability to create multiple interpretations of materials. To draw maximum benefit from students' use of these materi-

als, Thompson proposed that instructors must continually ask “What do I want my students to understand?” rather than “What do I want my students to do?”

The importance of looking at the total instructional environment is demonstrated in a study by Moyer (2002). She tracked how 10 teachers used manipulatives after they attended a 2-week summer institute with a Middle Grades Mathematics Kit that included base-10 blocks, color tiles, snap cubes, pattern blocks, fractions bars, and tangrams. The teachers made subtle distinctions between “real math” that used rules, procedures, and paper-and-pencil tasks and “fun math” that used the manipulatives. Unfortunately, the fun math typically was done at the end of the period or the end of the week and was not well integrated with the real math.

However, recent research inspired by theories of embodied cognition (Wilson, 2002) is discovering instructional approaches in which manipulatives enhance learning. An example is the study by Glenberg, Gutierrez, Levin, Japuntich, and Kaschak (2004). The instructional method required young readers to simulate actions described in the text by interacting with toy objects such as a horse, a tractor, and a barn in a text about a farm. Both actual manipulation and imagined manipulation greatly increased memory and comprehension of the text when compared to a control group that read the text twice.

Should Individual Differences in Verbal and Spatial Abilities Influence Instructional Design?

Mayer and Massa (2003) recently examined the hypothesis that some people are verbal learners and some people are visual learners by conducting a correlational and factor analysis of 14 cognitive measures related to the visualizer–verbalizer dimension. Factor analysis resulted in the discovery of four factors, two of which are spatial ability and learning preference. Spatial ability was measured by standardized tests such as card rotations and paper folding. The learning preferences test asked students to assume they needed help in understanding the scientific text that Mayer (2001) had used in previous research. Students indicated the strength of their preference for either a verbal help screen that defined terms such as water vapor, ice crystals, and freezing level or a visual help screen that provided a diagram to accompany the text.

A striking aspect of Mayer and Massa’s (2003) results is the lack of a correlation between spatial ability and learning preferences. Performance on the card rotations test correlated .02, .04, and $-.08$ with three variations of the learning preference test. The paper folding test correlated .13, .00, and .07 with the learning preference tests. Why are these correlations so low and what are the instructional implications?

There are several possible (compatible) reasons why the correlations are so low. One reason is that it is not intuitively clear which help screen should be selected by a student with much higher spatial ability than verbal ability. Should he or she select a visual help screen to build on his or her strengths

or a verbal help screen to compensate for his or her weaknesses? A second reason is that students who do well on a dynamic visual test such as rotating an object (to measure spatial ability) would not necessarily do well in generating a picture from a text and therefore benefit from a picture in the learning preferences test. As argued by Kosslyn (1994), imagery is not a unitary task. Rather, each of a number of different subsystems interact to determine performances across various imagery tasks. A third explanation is that instructional content varies even more than imagery tasks, making it difficult to predict what kind of help a student will need in a given situation. A student who knows the definitions of water vapor and ice crystal might not know the definitions of force and acceleration.

The instructional consequence of these low correlations, and their possible explanations, is that it is likely futile to try to assign students to different instructional conditions based on their aptitudes. A more promising approach would be to allow students to click on pop-up help screens when they need assistance. They could click on the term “water vapor” to receive a definition or click on a diagram whenever they needed a visual aid. A limitation of this approach, however, is that students do not always use help screens in a very effective manner (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003). Formative evaluations of a particular instructional design would therefore be extremely important to evaluate its effectiveness.

When practical, a better approach would be to use a variety of different formats to explain difficult concepts. This approach was taken in the Animation Tutor: Average speed module to explain the counterintuitive concept that the average speed of a round trip can not exceed twice the slower speed (Reed & Jazo, 2002). Learners viewed the asymptote of a graph, used the definition of speed as the ratio of total distance to total time, and studied the limit of an algebraic function. When asked which of these approaches was the most helpful for explaining the constraint on average speed, 10 students selected the algebraic approach, 8 students selected the definition-based approach, and 6 students selected the graphic approach. Ideally, of course, students should see the interconnections among all three representations.

CONCLUSION

One of the advantages of comparing the different architectures in Table 1 is that it illustrates the richness of the input and coding that occur during multimedia learning. Although I have limited the discussion of multimedia to verbal and visual input, I have tried to convey the many different kinds of coding that can result from this input. In addition to such general instructional objectives as increasing recall, reducing interference, minimizing cognitive load, and enhancing understanding, there are particular instructional issues raised by whether learning consists of associating pairs of words, re-

calling a chess board, integrating information in a text and diagram, understanding how a device works, constructing an equation for an algebra word problem, or manipulating manipulatives.

Comparing cognitive architectures for multimedia learning should help researchers, theorists, and instructional designers take advantage of both the similarities and differences of these different approaches. It should also help them provide answers to the challenging questions raised in the previous section. Tentative answers to these questions are that (a) there is general agreement among the different architectures, which differ in focus; (b) learners' integration of multiple codes is underspecified in the models; (c) animated instruction is not required when mental simulations are sufficient; (d) actions must be meaningful to be successful; and (e) multimodal instruction is superior to targeting modal-specific individual differences.

ACKNOWLEDGMENTS

Partial support for this work was provided by the National Science Foundation's Course, Curriculum, and Laboratory Improvement Program under grant DUE-9950746 (an animation-based tutor for algebra word problems). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

I thank anonymous reviewers for their helpful comments on previous versions of this article.

REFERENCES

- Ainsworth, S. (1999). The functions of multiple representations. *Computers and Education*, 33, 131-152.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73, 277-320.
- Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic-keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, 104, 126-133.
- Baddeley, A. (1992). Is working memory working? The fifteenth Bartlett lecture. *Quarterly Journal of Experimental Psychology*, 44A, 1-31.
- Baddeley, A. (2001). Is working memory still working? *American Psychologist*, 56, 851-864.
- Baddeley, A., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, 129, 126-145.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158-173.
- Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 17-90). Orlando, FL: Academic.
- Barsalou, L. (2003). Situated simulation in the human cognitive system. *Language and Cognitive Processes*, 18, 513-562.
- Barsalou, L. W., Solomon, K. O., & Wu, L. L. (1999). Perceptual simulation in conceptual tasks. In M. K. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, psychological, and typological issues in cognitive linguistics* (pp. 209-228). Amsterdam: John Benjamins.
- Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3, 472-517.
- Clements, D. H. (1999). "Concrete" manipulatives, concrete ideas. *Contemporary Issues in Early Childhood*, 1, 45-60.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Engelkamp, J. (1998). *Memory for actions*. Hove, UK: Psychology Press.
- Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science*, 26, 817-831.
- Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition*, 21, 115-124.
- Glenberg, A. M., Gutierrez, T., Levin, J., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, 96, 424-436.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558-565.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Koriat, A., & Pearlman-Avni, S. (2003). Memory organization of action events and its relationship to memory performance. *Journal of Experimental Psychology: General*, 132, 435-454.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Marschark, M., & Hunt, R. R. (1989). A reexamination of the role of imagery in learning and memory. *Journal of Experimental Psychology: Human Learning and Memory*, 15, 710-720.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge, UK: Cambridge University Press.
- Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology*, 95, 833-846.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, 43-52.
- Moyer, P. S. (2002). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics*, 47, 175-197.
- Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9, 329-389.
- Nathan, M. J., & Resnick, L. B. (1994). Less can be more: Unintelligent tutoring based on psychological theories and experimentation. In S. Vosniadou, E. De Corte, & H. Mandl (Eds.), *Technology-based learning environments* (pp. 183-192). Berlin: Springer-Verlag.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76, 241-263.
- Paivio, A. (1975). Coding distinctions and repetition effects in memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 179-214). Orlando, FL: Academic.
- Reed, S. K. (2005). From research to practice and back: The Animation Tutor project. *Educational Psychology Review*, 17, 55-83.
- Reed, S. K., & Jazo, L. (2002). Using multiple representations to improve conceptions of average speed. *Journal of Educational Computing Research*, 27, 147-166.
- Saariluoma, P. (1992). Visuospatial and articulatory interference in chess players' information intake. *Applied Cognitive Psychology*, 6, 77-89.
- Sanders, A. F., & Schroots, J. J. F. (1969). Cognitive categories and memory span: III. Effects of similarity on recall. *Quarterly Journal of Experimental Psychology*, 21, 21-28.
- Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14, 101-120.

- Steffens, M. C., Buchner, A., & Wender, K. F. (2003). Quite ordinary retrieval cues may determine free recall of actions. *Journal of Memory and Language, 48*, 399–415.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312.
- Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 215–266). San Diego, CA: Academic.
- Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science, 21*, 305–350.
- Thompson, P. W. (1994). Concrete materials and teaching for mathematical understanding. *Arithmetic Teacher, 40*, 556–558.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 1–16.
- Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language, 50*, 395–407.
- Watson, J. B. (1924). *Behaviorism*. New York: Norton.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636.
- Zwan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35–62). San Diego, CA: Academic.