3 Cognitive Theory of Multimedia Learning

Richard E. Mayer
University of California, Santa Barbara

Abstract

A fundamental hypothesis underlying research on multimedia learning is that multimedia instructional messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not so designed. The cognitive theory of multimedia learning is based on three cognitive science principles of learning: the human information processing system includes dual channels for visual/pictorial and auditory/verbal processing (i.e., dual-channel assumption), each channel has a limited capacity for processing (i.e., limited-capacity assumption), and active learning entails carrying out a coordinated set of cognitive processes during learning (i.e., active processing assumption). The cognitive theory of multimedia learning specifies five cognitive processes in multimedia learning: selecting relevant words from the presented text or narration, selecting relevant images from the presented graphics, organizing the selected words into a coherent verbal representation, organizing selected images into a coherent pictorial representation, and integrating the pictorial and verbal representations and prior knowledge. Three demands on the learner's cognitive capacity during learning are extraneous processing (which is not related to the instructional objective), essential processing (which is needed to mentally represent the essential material as presented), and generative processing (which is aimed at making sense of the material). Three instructional goals are to reduce extraneous processing (for extraneous overload situations), manage essential processing (for essential overload situations), and foster generative processing (for generative underuse situations). Multimedia instructional messages should be designed to guide appropriate cognitive processing during learning without overloading the learner's cognitive system.

The Case for Multimedia Learning

What Is the Rationale for a Theory of Multimedia Learning?

People learn more deeply from words and pictures than from words alone. This assertion – which can be called the *multimedia principle* – underlies

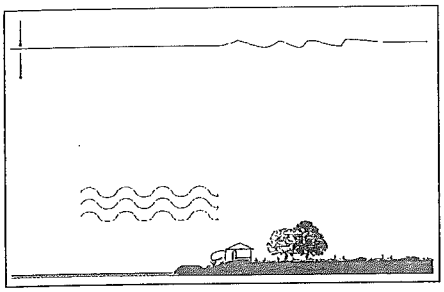
much of the interest in multimedia learning. For thousands of years, words have been the major format for instruction – including spoken words and, within the past few hundred years, printed words. Today, thanks to advances in computer and communication technologies, pictorial forms of instruction are becoming widely available, including dazzling computer-based graphics. However, simply adding pictures to words does not guarantee an improvement in learning – that is, all multimedia presentations are not equally effective. In this chapter I explore a theory aimed at understanding how to use words and pictures to improve human learning.

A fundamental hypothesis underlying research on multimedia learning is that multimedia instructional messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not so designed. For the past 25 years my colleagues and I at the University of California, Santa Barbara (UCSB) have been engaged in a sustained effort to construct an evidenced-based theory of multimedia learning that can guide the design of effective multimedia instructional messages (Mayer 2001, 2008, 2009; Mayer & Moreno, 2003).

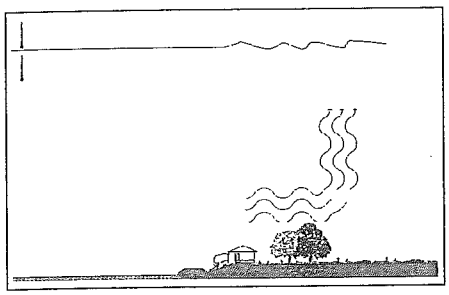
What Is a Multimedia Instructional Message?

A multimedia instructional message is a communication containing words and pictures intended to foster learning. The communication can be delivered using any medium, including paper (i.e., book-based communications) and computers (i.e., computer-based communications), or even face to face (i.e., face-to-face communications). Words can include printed words (such as you are now reading) or spoken words (such as in a narration); pictures can include static graphics – such as illustrations, charts, and photos – or dynamic graphics – such as animation and video clips. This definition is broad enough to include textbook chapters containing text and illustrations, online lessons containing animation and narration, interactive simulation games including on-screen text and graphics, and face-to-face slideshow presentations involving graphics and spoken words. For example, Figure 3.1 presents frames from a narrated animation on lightning formation, which we have studied in numerous experiments (Mayer, 2009).

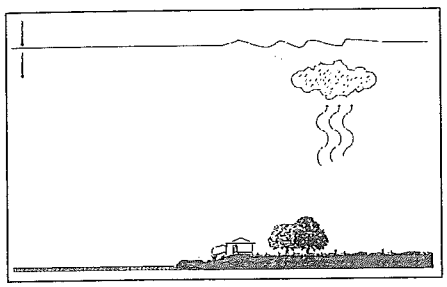
Learning can be measured by tests of retention (i.e., remembering the presented information) and transfer (i.e., being able to use the information to solve new problems), as described in Chapter 1. Our focus is on transfer test performance because we are mainly interested in how words and pictures can be used to promote understanding. In short, transfer tests can help tell us how well people understand what they have learned. We are particularly interested in the cognitive processes by which people construct meaningful learning outcomes from words and pictures.



"Cool moist air moves over a warmer surface and becomes heated."



"Warmed moist air near the earth's surface rises rapidly."



"As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud."

Figure 3.1. *Selected frames from a narrated animation on lightning formation.*

What Is the Role of a Theory of Learning in Multimedia Design?

Much of the work presented in this handbook is based on the premise that the design of multimedia instructional messages should be compatible with how people learn. In short, the design of multimedia instructional messages should be sensitive to what we know about how people process information. The cognitive theory of multimedia learning represents an attempt to accomplish this goal by describing how people learn from words and pictures, in a way that is consistent with empirical research evidence (e.g., Mayer, 2001, 2008, 2009; Mayer & Moreno, 2003) and consensus principles in cognitive science (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Mayer, 2011).

In building the cognitive theory of multimedia learning, my colleagues and I were guided by four criteria: *theoretical plausibility* – the theory is consistent with cognitive science principles of learning; *testability* – the theory yields predictions that can be tested in scientific research; *empirical plausibility* – the theory is consistent with empirical research evidence on multimedia learning; and *applicability* – the theory is relevant to educational needs for improving the design of multimedia instructional messages. In this chapter, I describe the cognitive theory of multimedia learning, which is intended to meet these criteria. In particular, I summarize three underlying assumptions of the theory derived from cognitive science; describe three memory stores, five cognitive processes, and five forms of representation in the theory; examine three demands on the learner's cognitive capacity during learning and three resulting goals for coping with them; and then provide a historical overview and a conclusion.

Three Assumptions of a Cognitive Theory of Multimedia Learning

Decisions about how to design a multimedia message always reflect an underlying conception of how people learn – even when the underlying theory of learning is not stated (Mayer, 1992). In short, the design of multimedia messages is influenced by the designer's conception of how the human mind works. For example, when a multimedia presentation consists of a screen overflowing with multicolored words and images – flashing and moving about – this reflects the designer's conception of human learning. The designer's underlying conception is that human learners possess a single-channel, unlimited-capacity, and passive processing system. First, by not taking advantage of auditory modes of presentation, this design is based on a single-channel assumption – all information enters the cognitive system in the same way regardless of its modality. It follows that it does not matter which modality is used to present information – such as presenting words as sounds or text – just as long as the information is presented. Second, by presenting so much information, this design is based on an unlimited-capacity assumption – humans can handle an unlimited amount of material. It follows that the designer's job is to present information to the learner. Third, by presenting many isolated pieces of information, this design is based on a passive processing assumption – humans act as if they were tape recorders, adding as much information to their memories as possible. It follows that learners do not need any guidance in organizing and making sense of the presented information.

What's wrong with this vision of learners as possessing a single-channel, unlimited-capacity, passive processing system? Current research in cognitive psychology paints a quite different picture of how the human mind works

Table 3.1. *Three assumptions of a cognitive theory of multimedia learning*

Assumption	Description	Related citations
Dual channels	Humans possess separate channels for processing visual and auditory information	Paivio (1986), Baddeley (1992)
Limited capacity	Humans are limited in the amount of information that can be processed in each channel at one time	Baddeley (1992), Chandler and Sweller (1991)
Active processing	Humans engage in active learning by attending to relevant incoming information, organizing selected information into coherent mental representations, and integrating mental representations with other knowledge	Mayer (1999), Wittrock (1989)

(Mayer, 2009, 2011). Thus, a difficulty with this commonsense conception of learning is that it conflicts with what is known about how people learn. In this section, I explore three assumptions underlying a cognitive theory of multimedia learning – *dual channels*, *limited capacity*, and *active processing*. These assumptions are summarized in Table 3.1.

Dual-Channel Assumption

The dual-channel assumption is that humans possess separate information processing channels for visually/spatially represented material and auditorily/verbally represented material. The relevance of the dual-channel assumption to the cognitive theory of multimedia learning lies in the proposal that the human information processing system contains an auditory/verbal channel and a visual/pictorial channel. When information is presented to the eyes (such as illustrations, animations, video, or on-screen text), humans begin by processing that information in the visual channel; when information is presented to the ears (such as narration or nonverbal sounds), humans begin by processing that information in the auditory channel. The concept of separate information processing channels has a long history in cognitive psychology and currently is most closely associated with Paivio's dual-coding theory (Clark & Paivio, 1991; Paivio, 1986, 2006) and Baddeley's model of working memory (Baddeley, 1999; Baddeley, Eysenck, & Anderson, 2009).

What is processed in each channel? There are two ways of conceptualizing the differences between the two channels – one based on *representation modes* and the other based on *sensory modalities*. The representation-mode approach focuses on whether the presented stimulus is verbal (such as spoken or printed words) or nonverbal (such as pictures, video, animation, or background sounds). According to the representation-mode approach, one channel processes verbal material and the other channel processes pictorial

material and nonverbal sounds. This conceptualization is most consistent with Paivio's (1986, 2006) distinction between verbal and nonverbal systems.

In contrast, the sensory-modality approach focuses on whether learners initially process the presented materials through their eyes (such as for pictures, video, animation, or printed words) or ears (such as for spoken words or background sounds). According to the sensory-modality approach, one channel processes visually represented material and the other channel processes auditorily represented material. This conceptualization is most consistent with Baddeley's (1999; Baddeley, Eysenck, & Anderson, 2009) distinction between the visuospatial sketchpad and the phonological loop.

Whereas the representation-mode approach focuses on the format of the stimulus (i.e., verbal or nonverbal), the sensory-modality approach focuses on the sensory modality of the stimulus (i.e., auditory or visual). The major difference concerning multimedia learning rests in the processing of printed words (i.e., on-screen text) and background sounds. On-screen text is initially processed in the verbal channel in the representation-mode approach but in the visual channel in the sensory-modality approach; background sounds, including nonverbal music, are initially processed in the nonverbal channel in the representation-mode approach but in the auditory channel in the sensory-modality approach.

For purposes of the cognitive theory of multimedia learning, I have opted for a compromise in which I use the sensory-modality approach to distinguish between visually presented material (such as pictures, animations, video, and on-screen text) and auditorily presented material (such as narration and background sounds), as well as a representation-mode approach to distinguish between the construction of pictorially based and verbally based models in working memory. However, additional research is necessary to clarify the nature of the differences between the two channels and the implications for learning and instruction.

What is the relation between the channels? Although information enters the human information system via one channel, learners may be able to convert the representation for processing in the other channel. When learners are able to devote adequate cognitive resources to the task, it is possible for information originally presented to one channel to also be represented in the other channel. For example, on-screen text may initially be processed in the visual channel because it is presented to the eyes, but an experienced reader may be able to mentally convert images into sounds, which are processed through the auditory channel. Similarly, an illustration of an object or event such as a cloud rising above the freezing level may initially be processed in the visual channel, but the learner may also be able to mentally construct the corresponding verbal description in the auditory channel. Conversely, a narration describing some event such as "the cloud rises above the freezing level" may initially be processed in the auditory channel because it is presented to

the ears, but the learner may also form a corresponding mental image that is processed in the visual channel. Cross-channel representations of the same stimulus play an important role in Paivio's (1986, 2006) dual-coding theory.

Limited-Capacity Assumption

The second assumption is that humans are limited in the amount of information that can be processed in each channel at one time. When an illustration or animation is presented, the learner is able to hold only a few images in the visual channel of working memory at any one time, reflecting portions of the presented material rather than an exact copy of the presented material. For example, if an illustration or animation of a tire pump is presented, the learner may be able to focus on building mental images of the handle going down, the inlet valve opening, and air moving into the cylinder. When a narration is presented, the learner is able to hold only a few words in the verbal channel of working memory at any one time, reflecting portions of the presented text rather than a verbatim recording. For example, if the spoken text is "When the handle is pushed down, the piston moves down, the inlet valve opens, the outlet valve closes, and air enters the bottom of cylinder," the learner may be able to hold the following verbal representations in auditory working memory: "handle goes up," "inlet valve opens," and "air enters cylinder." The conception of limited capacity in consciousness has a long history in psychology, and some modern examples are Baddeley's (1999; Baddeley, Eysenck, & Anderson, 2009; see also Chapter 25) theory of working memory and Sweller's (1999; Sweller, Ayres, & Kalyuga, 2011; see also Chapter 2) cognitive load theory.

What are the limits on cognitive capacity? If we assume that each channel has limited processing capacity, it is important to know just how much information can be processed in each channel. The classic way to measure someone's cognitive capacity is to give a memory span test (Miller, 1956; see also Mayer, 2011), although more recent advancements include the OSpan and RSpan tests, as described in Chapter 25. Although there are individual differences, on average, memory span is fairly small – approximately five to seven chunks.

With practice, of course, people can learn techniques for chunking the elements in the list, such as grouping the seven digits 8–7–5–3–9–6–4 into three chunks, 875–39–64 (e.g., "eight seven five" pause "three nine" pause "six four"). In this way, the cognitive capacity remains the same – five to seven chunks – but more elements can be remembered within each chunk (Mayer, 2011).

How are limited cognitive resources allocated? The constraints on our processing capacity force us to make decisions about which pieces of incoming information to pay attention to, the degree to which we should build connections among the selected pieces of information, and the degree to which

we should build connections between selected pieces of information and our existing knowledge. *Metacognitive strategies* are techniques for allocating, monitoring, coordinating, and adjusting these limited cognitive resources. These strategies are at the heart of what Baddeley (1999; Baddeley, Eysenck, & Anderson, 2009) calls the *central executive* – the system that controls the allocation of cognitive resources – and play a central role in modern theories of metacognition (Hacker, Dunlosky, & Graesser, 2009).

Active Processing Assumption

The third assumption is that humans actively engage in cognitive processing in order to construct a coherent mental representation of their experiences. These active cognitive processes include paying attention to relevant incoming information, organizing incoming information into a coherent cognitive structure, and integrating incoming information with other knowledge. In short, humans are active processors who seek to make sense of multimedia presentations. This view of humans as active processors conflicts with a common view of humans as passive processors who seek to add as much information as possible to memory, that is, as if they were tape recorders filing copies of their experiences in memory to be retrieved later.

What are the major ways that knowledge can be structured? Active learning occurs when a learner applies cognitive processes to incoming material – processes that are intended to help the learner make sense of the material. The desired outcome of active cognitive processing is the construction of a coherent mental representation, so active learning can be viewed as a process of model building. A *mental model* (*or knowledge structure*) represents the key parts of the presented material and their relations. For example, in a multimedia presentation of how lightning storms develop, the learner may attempt to build a cause-and-effect system in which a change in one part of the system causes a change in another part. In a lesson comparing and contrasting two theories, construction of a mental model involves building a sort of matrix structure that compares the two theories along several dimensions.

If the outcome of active learning is the construction of a coherent mental representation, it is useful to explore some of the typical ways that knowledge can be structured. Some basic knowledge structures include *process*, *comparison*, *generalization*, *enumeration*, and *classification* (Chambliss & Calfee, 1998; Cook & Mayer, 1988). Process structures can be represented as cause-and-effect chains and consist of explanations of how some system works. An example is an explanation of how the human ear works. Comparison structures can be represented as matrices and consist of comparisons among two or more elements along several dimensions. An example is a comparison between how two competing theories of learning view the role of the learner, the role of the teacher, and useful types of instructional methods.

Table 3.2. *Three cognitive processes required for active learning*

Process	Description
Selecting	Attending to relevant material in the presented lesson for transfer to working memory
Organizing	Mentally organizing selected information into a coherent cognitive structure in working memory
Integrating	Connecting cognitive structures with each other and with relevant prior knowledge activated from long-term memory

Generalization structures can be represented as a branching tree and consist of a main idea with subordinate supporting details. An example is an essay in support of lowering the voting age. Enumeration structures can be represented as lists and consist of a collection of items. An example is the names of principles of multimedia learning listed in this handbook. Classification structures can be represented as hierarchies and consist of sets and subsets. An example is a biological classification system for sea animals.

Understanding a multimedia message often involves constructing one or more of these kinds of knowledge structures. This assumption suggests two important implications for multimedia design: (1) the presented material should have a coherent structure, and (2) the message should provide guidance to the learner on how to build the structure. If the material lacks an underlying coherent structure – for example, if the material is mainly a collection of isolated facts – the learner's model-building efforts will be fruitless. If the message lacks guidance on how to structure the presented material, the learner's model-building efforts may be overwhelmed. Multimedia design can be conceptualized as an attempt to assist learners in their model-building efforts.

What are the cognitive processes involved in active learning? Table 3.2 summarizes three cognitive processes that are essential for active learning: selecting relevant material, organizing selected material, and integrating selected material with existing knowledge (Mayer, 2009; Wittrock, 1989). Selecting relevant material occurs when a learner pays attention to appropriate words and images in the presented material. This process involves bringing material from the outside into the working memory component of the cognitive system. Organizing selected material involves building structural relations among the elements – such as one of the five kinds of structures described in the preceding section. This process takes place within the working memory component of the cognitive system. Integrating selected material with existing knowledge involves building connections between incoming material and relevant portions of prior knowledge. This process involves activating knowledge in long-term memory and bringing it into working memory. For example, in a multimedia message on the cause of lightning, learners must pay attention to certain words and images, arrange them into a

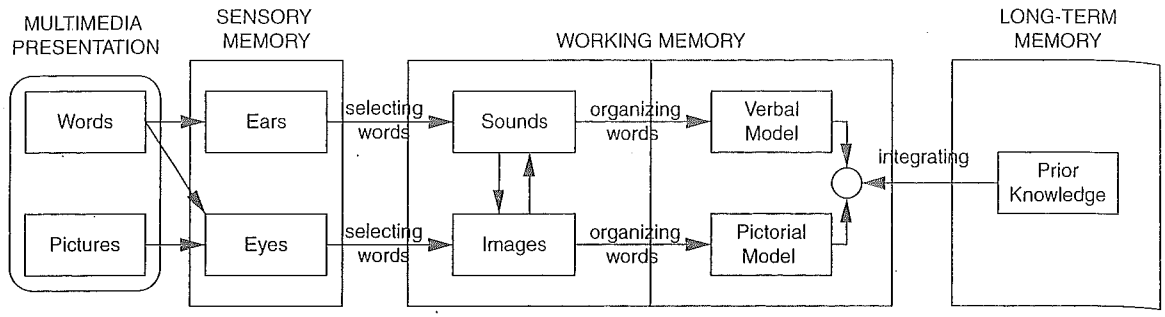


Figure 3.2. *Cognitive theory of multimedia learning.*

cause-and-effect chain, and relate the steps to prior knowledge such as the principle that hot air rises.

In sum, the implicit theory of learning underlying some multimedia messages is that learning is a single-channel, unlimited-capacity, passive processing activity. In contrast, I offer a cognitive theory of multimedia learning that is based on three basic assumptions about how the human mind works – namely, that the human mind is a dual-channel, limited-capacity, active processing system.

Three Memory Stores in the Cognitive Theory of Multimedia Learning

Figure 3.2 presents a cognitive model of multimedia learning intended to represent the human information processing system. The boxes represent memory stores, including sensory memory, working memory, and long-term memory, and the arrows represent the cognitive processes of selecting, organizing, and integrating. The top row represents the verbal channel and the bottom row represents the visual channel.

Table 3.3 summarizes the characteristics of the three memory stores in the cognitive theory of multimedia learning. Pictures and words come in from the outside world as a multimedia presentation (indicated on the left side of the figure) and enter sensory memory through the eyes and ears (indicated in the "Sensory Memory" box). Sensory memory allows for pictures and printed text to be held as exact visual images for a very brief time period in a visual sensory memory (at the top) and for spoken words and other sounds to be held as exact auditory images for a very brief time period in an auditory sensory memory (at the bottom). The arrow from "Pictures" to "Eyes" corresponds to a picture being registered in the visual sensory memory, the arrow from "Words" to "Ears" corresponds to spoken text being registered in the auditory sensory memory, and the arrow from "Words" to "Eyes" corresponds to printed text being registered in the visual sensory memory.

Table 3.3. *Three memory stores in the cognitive theory of multimedia learning*

Memory store	Description	Capacity	Duration	Format
Sensory memory	Briefly holds sensory copies of incoming words and pictures	Unlimited	Very brief	Visual or auditory sensory images
Working memory	Allows for manipulating selected incoming information	Limited	Short	Verbal and pictorial representations
Long-term memory	Permanently stores organized knowledge	Unlimited	Permanent	Knowledge

The central work of multimedia learning takes place in working memory, so let's focus on the "Working Memory" box in Figure 3.2. Working memory is used for temporally holding and manipulating knowledge in active consciousness. For example, in reading this sentence you may be able to actively concentrate on only some of the words at one time, or in looking at Figure 3.2 you may be able to hold the images of only some of the boxes and arrows in your mind at one time. This kind of processing – namely, processing that involves conscious attention – takes place in working memory. The left side of the "Working Memory" box represents the raw material that comes into working memory – visual images of pictures and sound images of words – so it is based on the two sensory modalities that I call visual and auditory; in contrast, the right side of the "Working Memory" box represents the knowledge constructed in working memory – pictorial and verbal models and links between them – so it is based on the two representation modes that I call pictorial and verbal. I use the term *pictorial model* to refer to spatial representations rather than visual images. The arrow from "Sounds" to "Images" represents the mental conversion of a sound (such as the spoken word "cat") into a visual image (such as an image of a cat) – that is, when you hear the word "cat" you might also form a mental image of a cat; the arrow from "Images" to "Sounds" represents the mental conversion of a visual image (such as a mental picture of a cat) into a sound (such as the sound of the word "cat") – that is, you mentally hear the word "cat" when you see a picture of one.

Finally, the box on the right is labeled "Long-Term Memory" and corresponds to the learner's storehouse of knowledge. Unlike working memory, long-term memory can hold large amounts of knowledge over long periods

of time, but to actively think about material in long-term memory it must be brought into working memory (as indicated by the arrow from "Long-Term Memory" to "Working Memory").

The major cognitive processing required for multimedia learning is represented by the arrows in Figure 3.2 labeled "selecting images," "selecting words," "organizing images," "organizing words," and "integrating" – which are described in the next section.

Five Processes in the Cognitive Theory of Multimedia Learning

For meaningful learning to occur in a multimedia environment, the learner must engage in five cognitive processes, indicated by the arrows in Figure 3.2: (1) selecting relevant words for processing in verbal working memory, (2) selecting relevant images for processing in visual working memory, (3) organizing selected words into a verbal model, (4) organizing selected images into a pictorial model, and (5) integrating the verbal and pictorial representations with each other and with relevant prior knowledge activated from long-term memory. The five cognitive processes in multimedia learning are summarized in Table 3.4. Although I present these processes as a list, they do not necessarily occur in linear order, so a learner might move from process to process in many different ways. Successful multimedia learning requires that the learner coordinate and monitor these five processes.

Selecting Relevant Words

The first labeled step in Figure 3.2 involves a change in knowledge representation from the external presentation of spoken words (such as a computer-

Table 3.4. *Five cognitive processes in the cognitive theory of multimedia learning*

Process	Description
Selecting words	Learner pays attention to relevant words in a multimedia message to create sounds in working memory
Selecting images	Learner pays attention to relevant pictures in a multimedia message to create images in working memory
Organizing words	Learner builds connections among selected words to create a coherent verbal model in working memory
Organizing images	Learner builds connections among selected images to create a coherent pictorial model in working memory
Integrating	Learner builds connections between verbal and pictorial models and with prior knowledge

generated narration) to a sensory representation of sounds to an internal working memory representation of word sounds (such as some of the words in the narration). The input for this step is a spoken verbal message – that is, the spoken words in the presented portion of the multimedia message. The output for this step is a word sound base (called *sounds* in Figure 3.2) – that is, a mental representation in the learner's verbal working memory of selected words or phrases.

The cognitive process mediating this change is called *selecting relevant words* and involves paying attention to some of the words that are presented in the multimedia message as they pass through auditory sensory memory. If the words are presented as speech, this process begins in the auditory channel (as indicated by the arrows from "Words" to "Ears" to "Sounds"). However, if the words are presented as on-screen text or printed text, this process begins in the visual channel (as indicated by the arrow from "Words" to "Eyes") and later may move to the auditory channel if the learner mentally articulates the printed words (as indicated by the arrow from "Images" to "Sounds" in the left side of working memory). The need for selecting only part of the presented message occurs because of capacity limitations in each channel of the cognitive system. If the capacity were unlimited, there would be no need to focus attention on only part of the verbal message. Finally, the selection of words is not arbitrary; the learner must determine which words are most relevant – an activity that is consistent with the view of the learner as an active sense maker.

For example, in the lightning lesson partially shown in Figure 3.1, one segment of the multimedia presentation contains the words "Cool moist air moves over a warmer surface and becomes heated," the next segment contains the words "Warmed moist air near the earth's surface rises rapidly," and the next segment has the words "As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud." When a learner engages in the selection process, the result may be that some of the words are represented in verbal working memory – such as, "Cool air becomes heated, rises, forms a cloud."

Selecting Relevant Images

The second step involves a change in knowledge representation from the external presentation of pictures (such as an animation segment or an illustration) to a sensory representation of unanalyzed visual images to an internal representation in working memory (such as a visual image of part of the animation or illustration). The input for this step is a pictorial portion of a multimedia message that is held briefly in visual sensory memory. The output for this step is a visual image base – a mental representation in the learner's working memory of selected images.

The cognitive process underlying this change – *selecting relevant images* – involves paying attention to part of the animation or illustrations presented in

the multimedia message. This process begins in the visual channel, but it is possible to convert part of it to the auditory channel (such as by mentally narrating an ongoing animation). The need to select only part of the presented pictorial material arises from the limited processing capacity of the cognitive system. It is not possible to process all parts of a complex illustration or animation simultaneously, so learners must focus on only part of the incoming pictorial material at a time. Finally, the selection process for images – like the selection process for words – is not arbitrary because the learner must judge which images are most relevant for making sense of the multimedia presentation.

In the lightning lesson, for example, one segment of the animation shows blue colored arrows – representing cool air – moving over a heated land surface that contains a house and trees; another segment shows the arrows turning red and traveling upward above a tree; and a third segment shows the arrows changing into a cloud with lots of dots inside. In selecting relevant images, the learner may compress all this into images of a blue arrow pointing rightward, a red arrow pointing upward, and a cloud; details such as the house and tree on the surface, the wavy form of the arrows, and the dots in the cloud are lost.

Organizing Selected Words

Once the learner has formed a word sound base from the incoming words of a segment of the multimedia message, the next step is to organize the words into a coherent representation – a knowledge structure that I call a *verbal model*. The input for this step is the word sound base – the word sounds selected from the incoming verbal message – and the output for this step is a verbal model – a coherent (or structured) representation in the learner's working memory of the selected words or phrases.

The cognitive process involved in this change is *organizing selected words*, in which the learner builds connections among pieces of verbal knowledge. This process is most likely to occur in the auditory channel and is subject to the same capacity limitations that affect the selection process. Learners do not have unlimited capacity to build all possible connections so they must focus on building a simple structure. The organizing process is not arbitrary, but rather reflects an effort at sense making – such as the construction of a cause-and-effect chain.

For example, in the lightning lesson partially shown in Figure 3.1, the learner may build causal connections between the selected verbal components: "First: cool air is heated; second: it rises; third: it forms a cloud." In mentally building a causal chain, the learner is organizing the selected words.

Organizing Selected Images

The process for organizing images parallels that for selecting words. Once the learner has formed an image base from the incoming pictures of a segment

of the multimedia message, the next step is to organize the images into a coherent representation – a knowledge structure that I call a *pictorial model*. The input for this step is the visual image base – the images selected from the incoming pictorial message – and the output for this step is a pictorial model – a structured spatial representation in the learner's working memory based on the selected images.

This change from images to pictorial model requires the application of a cognitive process that I call *organizing selected images*. In this process, the learner builds connections among pieces of pictorial knowledge. This process occurs in the visual channel, which is subject to the same capacity limitations that affect the selection process. Learners lack the capacity to build all possible connections among images in their working memory, but rather must focus on building a simple set of connections. As in the process of organizing words, the process of organizing images is not arbitrary. Rather, it reflects an effort to build a simple structure that makes sense to the learner – such as a cause-and-effect chain.

For example, in the lightning lesson, the learner may build causal connections between the selected images: the rightward-moving blue arrow turns into a rising red arrow, which turns into a cloud. In short, the learner builds causal links in which the first event leads to the second and so on.

Integrating Word-Based and Picture-Based Representations

Perhaps the most crucial step in multimedia learning involves making connections between word-based and picture-based representations. This step involves a change from having two separate representations – a verbal model and a pictorial model – to having an integrated representation in which corresponding elements and relations from one model are mapped onto the other. The input for this step is the pictorial model and the verbal model that the learner has constructed so far, and the output is an integrated model, which is based on connecting the two representations. In addition, the integrated model includes connections with relevant prior knowledge.

I refer to this cognitive process as *integrating words and images* because it involves building connections between corresponding portions of the pictorial and verbal models as well as with relevant knowledge from long-term memory. This process occurs in visual and verbal working memory and involves the coordination between them. This is an extremely demanding process that requires the efficient use of cognitive capacity. The process reflects the epitome of sense making because the learner must focus on the underlying structure of the visual and verbal representations. The learner also can use prior knowledge activated from long-term memory to help coordinate the integration process, as indicated by the arrow from long-term memory to working memory.

For example, in the lightning lesson, the learner must see the connection between the verbal chain – "First, cool air is heated; second, it rises; third, it

forms a cloud" – and the pictorial chain – the blue arrow followed by the red arrow followed by the cloud shape. In addition, the learner can apply prior knowledge to the transition from the first to the second event by remembering that hot air rises.

Each of the five processes in multimedia learning is likely to occur many times throughout a multimedia presentation. The processes are applied segment by segment rather than to the message as a whole. For example, in processing the lightning lesson, learners do not first select all relevant words and images from the entire passage, then organize them into verbal and pictorial models of the entire passage, and then connect the completed models with one another at the very end. Rather, learners carry out this procedure on small segments: they select relevant words and images from the first sentence of the narration and the first few seconds of the animation; they organize and integrate them; and then this set of processes is repeated for the next segment, and so on. Schnotz and Bannert's (2003; see also Chapter 4) integrated model of text and picture comprehension also addresses the issue of how learners integrate words and pictures.

Finally, another process (not shown in Figure 3.2 or Table 3.4) is *encoding*, which involves an arrow from working memory to long-term memory, signifying the transfer of the constructed representation from working memory to long-term memory for permanent storage within the learner's organized knowledge base.

Five Forms of Representation

As you can see in Figure 3.2, there are five forms of representation for words and pictures, reflecting their stage of processing. To the far left, we begin with *words and pictures in the multimedia presentation* – that is, the stimuli that are presented to the learner. In the case of the lightning message shown in Figure 3.1, the words are the spoken words presented through the computer's speakers and the pictures are the frames of the animation presented on the computer screen. Second, as the presented words and pictures impinge on the learner's ears and eyes, the next form of representation is *acoustic representations (or sounds) and visual representations (or images) in sensory memory.* The sensory representations fade rapidly, unless the learner pays attention to them. Third, when the learner selects some of the words and images for further processing in working memory, the next form of representation is *sounds and images in working memory*. These are the building blocks of knowledge construction – including key phrases such as "warmed air rises" and key images such as red arrows moving upward. The fourth form of representation results from the learner's construction of a *verbal model and pictorial model in working memory*. Here the learner has organized the material into coherent verbal and spatial representations, and also has

Table 3.5. *Five forms of representation in the cognitive theory of multimedia learning*

Type of knowledge	Location	Example
Words and pictures	Multimedia presentation	Sound waves from computer speaker: "Cool moist air ..."; pixel patterns on the computer screen showing a wavy blue arrow
Acoustic and iconic representations	Sensory memory	Received sounds in learner's ears: "Cool moist air ..."; received image in learner's eyes corresponding to wavy blue arrow
Sounds and images	Working memory	Selected sounds: "Cool moist air moves"; selected images: wavy blue line moving rightward
Verbal and pictorial models	Working memory	Mental model of cloud formation
Prior knowledge	Long-term memory	Schema of differences in air pressure

mentally integrated them. The pictorial model should be considered a schematic spatial representation rather than a sensory-like visual image. Finally, the fifth form of representation is *knowledge in long-term memory*, which the learner uses for guiding the process of knowledge construction in working memory. Sweller (1999) refers to this knowledge as *schemas*. As new knowledge is constructed in working memory, it may be stored in long-term memory as prior knowledge to be used in supporting new learning. The five forms of representation are summarized in Table 3.5.

Three Kinds of Demands on Cognitive Capacity

The challenge for instructional design is to guide the learner's appropriate cognitive processing during learning without overloading the learner's working memory capacity. Table 3.6 summarizes three kinds of demands on the learner's information processing system during learning: extraneous processing, essential processing, and generative processing.

Extraneous processing refers to cognitive processing that does not support the instructional goal and is caused by poor instructional design. For example, when a figure is printed on one page and the words describing the figure are printed on another page, a learner may have to scan back and forth, resulting in extraneous processing that wastes precious cognitive capacity. Extraneous processing does not result in any useful knowledge being constructed in the learner's working memory. Extraneous processing is analogous to *extraneous cognitive load* in cognitive load theory, as described in Chapter 2.

Table 3.6. *Three demands on cognitive capacity during multimedia learning*

Name	Description	Caused by	Learning processes	Example
Extraneous processing	Cognitive processing that is not related to the instructional goal	Poor instructional design	None	Focusing on irrelevant pictures
Essential processing	Cognitive processing to represent the essential presented material in working memory	Complexity of the material	Selecting	Memorizing the description of essential processing
Generative processing	Cognitive processing aimed at making sense of the material	Motivation to learn	Organizing and integrating	Explaining generative processing in one's own words

Essential processing refers to cognitive processing aimed at mentally representing the presented material in working memory and is caused by the complexity of the material. For example, less essential processing is required to mentally represent the definition of working memory than is required to mentally represent the information processing system summarized in Figure 3.2. Essential processing involves selecting relevant information from the presentation and organizing it as presented. Thus, essential processing results in the construction of verbal and pictorial representations in working memory that correspond to the presented material, analogous to a *text-base* in Kintsch's (1998) construction-integration theory of text processing. Essential processing is analogous to *intrinsic cognitive load* in cognitive load theory, as described in Chapter 2.

Generative processing refers to cognitive processing aimed at making sense of the presented material and is caused by the learner's motivation to learn. For example, when the material is presented by a likable instructor, the learner may exert more effort to understand what the instructor is presenting. Generative processing involves reorganizing the incoming information and integrating it with relevant prior knowledge. Thus, generative processing results in the construction of an integrated mental model, analogous to a *situation model* in Kintsch's (1998) construction-integration theory of text processing. Generative processing is analogous to *germane cognitive load* in cognitive load theory, as described in Chapter 2. Both generative and essential processes are directed at the instructional goal.

Each of the key concepts – cognitive capacity, extraneous processing, essential processing, and generative processing – is relative to the learner and the learner's interaction with the instructional situation. For example, learners differ in terms of their working memory capacity (as explored in

Chapter 25), which affects their ability to handle each of the three kinds of demands on cognitive capacity. Learners differ in their cognitive and metacognitive strategies for engaging in generative processing and essential processing. They differ in terms of their prior knowledge that can help them handle the extraneous processing caused by poorly designed instructional situations or guide their essential and generative processing of familiar material. For example, individual differences in prior knowledge are an important consideration in the instructional design of multimedia instruction (see Chapter 24). Thus, the identical multimedia lesson may be overloading for one learner and not be overloading for another because of differences in the capacities, knowledge, skills, and beliefs (e.g., beliefs about how learning works) that learners bring to the learning situation.

The learner has a limited amount of cognitive capacity to process information in each channel in working memory during learning, so capacity that is used for extraneous processing cannot be used for essential and generative processing. In short, consistent with cognitive load theory (Plass, Moreno, & Brunken, 2010; Sweller, Ayres, & Kalyuga, 2011; also see Chapter 2), the sum of extraneous processing plus essential processing plus generative processing cannot exceed the learner's cognitive capacity. Given that the learner's cognitive capacity is limited and the three demands on cognitive capacity are additive, if the learner increases one kind of processing then another one must be decreased. The instructional implications of this triarchic model of cognitive processing demands are explored in the next section.

Three Learning Scenarios

Figure 3.3 summarizes three learning scenarios based on the triarchic model of cognitive processing demands. First, in the top frame, consider what happens when the instructional message is so poorly designed that the learner is forced to expend large amounts of processing capacity on extraneous processing, thereby leaving insufficient capacity for essential and generative processing. This scenario, which can be called *extraneous overload*, can be addressed by devising instructional methods aimed at reducing extraneous processing. Examples of techniques aimed at reducing extraneous processing include the coherence principle, signaling principle, redundancy principle, spatial contiguity principle (or split-attention principle), and temporal contiguity principle, as described in Chapters 8, 10, 11, and 13. The goal of these instructional techniques, which are summarized in Table 3.7, is to reduce extraneous processing so that available cognitive capacity can be used for essential and generative processing.

Next, consider what might happen when the learner is given an instructional message that is well designed so it does not create high levels of extraneous processing. The second frame in Figure 3.3 represents the *essential*

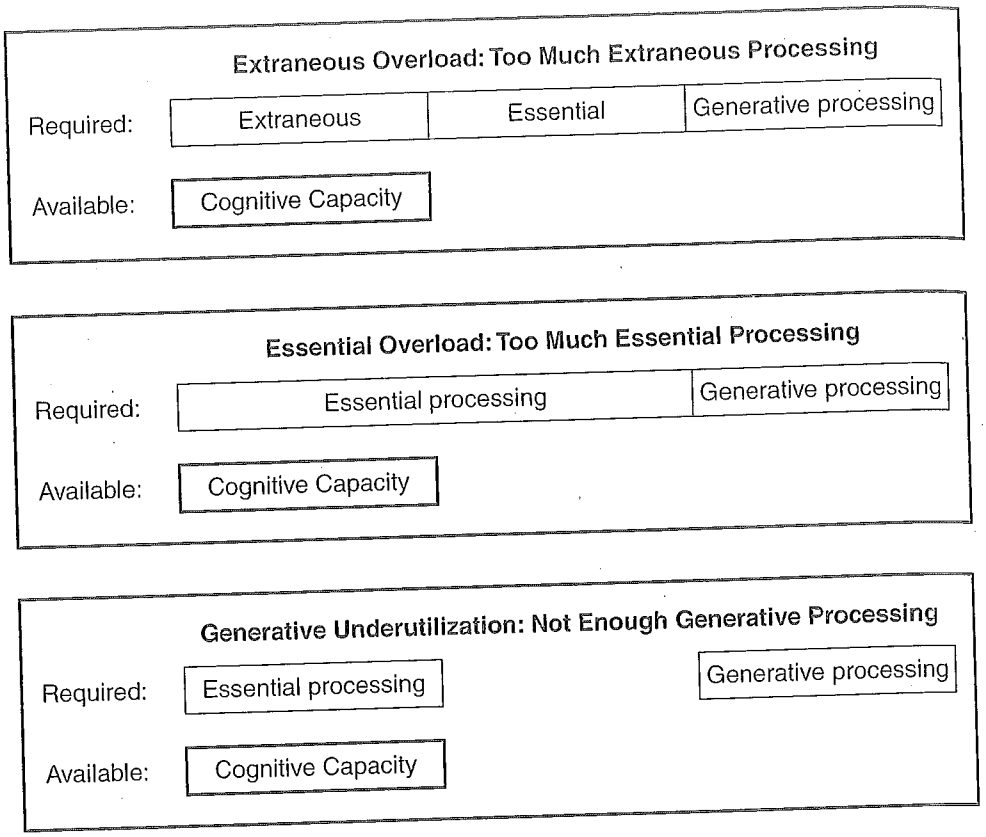


Figure 3.3. *Three learning scenarios.*

overload scenario, in which the material is so complicated that the learner does not have enough cognitive capacity to mentally represent it as presented. Essential processing is needed to mentally represent the to-be-learned material, so it is not appropriate to seek to reduce essential processing. In this case, a reasonable instructional goal is to *manage essential processing*. As summarized in Table 3.7, some instructional techniques aimed at managing essential processing are the segmenting principle, pre-training principle, and modality principle, as described in Chapters 9 and 12.

Finally, suppose that the learner receives multimedia instruction that is designed to minimize extraneous processing and manage essential processing, so there is cognitive capacity available for generative processing. The third frame in Figure 3.3 represents the *generative underutilization* scenario, in which the learner has cognitive capacity available to engage in generative processing but does not exert the effort to do so. The solution to this instructional problem is to *foster generative processing*, as summarized in Table 3.7. In short, the goal is to motivate learners to exert and maintain effort to make sense of the material at a sufficient level of intensity. Some instructional design techniques aimed at fostering generative processing include the multimedia principle, personalization principle, voice principle, and embodiment principle (as explored in Chapters 7 and 13). Some learning strategies aimed at priming generative processing during learning include the guided discovery

Table 3.7. *Three instructional goals in multimedia learning*

Goal	Representative technique	Description of technique	Chapter
Minimize extraneous processing	Coherence principle	Eliminate extraneous material	12
	Signaling principle	Highlight essential material	11, 12
	Redundancy principle	Do not add printed text to spoken text	10, 12
	Spatial contiguity principle	Place printed text near corresponding graphic	8, 12
	Temporal contiguity principle	Present narration and corresponding graphic simultaneously	12
	Segmenting principle	Break presentation into parts	13
Manage essential processing	Pre-training principle	Describe names and characteristics of key elements before the lesson	13
	Modality principle	Use spoken rather than printed text	9, 13
	Multimedia principle	Use words and pictures rather than words alone	7
Foster generative processing	Personalization principle	Put words in conversational style	14
	Voice principle	Use human voice for spoken words	14
	Embodiment principle	Give on-screen characters humanlike gestures	14
	Guided discovery principle	Provide hints and feedback as learner solves problems	15
	Self-explanation principle	Ask learners to explain a lesson to themselves	17
	Drawing principle	Ask learners to make drawings for the lesson	18

principle (Chapter 15), self-explanation principle (Chapter 17), and drawing principle (Chapter 18).

In summary, the cognitive theory of multimedia learning suggests three primary goals of instructional design: reduce extraneous processing, manage essential processing, and foster generative processing. The instructional techniques described in this handbook can be analyzed in terms of the kind of instructional goals they seek to address – helping students reduce their extraneous processing during learning (which was the original focus of much research in multimedia learning), helping students manage their essential

processing during learning (in which the modality principle has enjoyed the most attention), or helping students engage in generative processing during learning (which is a newer and less researched domain).

Historical Overview

The Past: Evolution of the Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning has evolved within the body of research papers and books produced by my colleagues and me at UCSB during the past 25 years. Although the name has changed over the years, the underlying elements of the theory – that is, dual channels, limited capacity, and active processing – have remained constant. Some names used early in the research program – such as "model of meaningful learning" (Mayer, 1989) and "cognitive conditions for effective illustrations" (Mayer & Gallini, 1990) – emphasized the active processing element; other names used later – such as "dual-coding model" (Mayer & Anderson, 1991, 1992) and "dual-processing model of multimedia learning" (Mayer & Moreno, 1998; Mayer, Moreno, Boire, & Vagge, 1999) – emphasized the dual-channel element; and yet other names – such as "generative theory" (Mayer, Steinhoff, Bower, & Mars, 1995) and "generative theory of multimedia learning" (Mayer, 1997; Plass, Chun, Mayer & Leutner, 1998) – emphasized all three elements. The current name, "cognitive theory of multimedia learning," was used in Mayer, Bove, Bryman, Mars, and Tapangco (1996), Moreno and Mayer (2000), and Mayer, Heiser, and Lonn (1991) and was selected for use in major reviews (Mayer, 2001, 2008, 2009; Mayer & Moreno, 2003) as well as the previous edition of *The Cambridge Handbook of Multimedia Learning* (Mayer, 2005).

An early predecessor of the flow chart representation shown in Figure 3.2 was a dual-coding model shown in Mayer and Sims (1994, fig. 1), which contained the same two channels and three of the same five cognitive processes but lacked two of the cognitive processes and sensory memory. Mayer, Steinhoff, Bower, and Mars (1995, fig. 1) and Mayer (1997, fig. 3) presented an intermediate version that was almost identical to the flow chart shown in Figure 3.2 except that it lacked long-term memory and sensory memory. Finally, the current version of the flow chart appeared in Mayer, Heiser, and Lonn (2001) and was reproduced in subsequent reviews (Mayer, 2001, fig. 3.2; 2002, fig. 7; 2003, fig. 2; 2005, fig. 3.2; 2009, fig. 3.1 Mayer & Moreno, 2003, fig. 1). Thus, the model has developed by the addition of components – both cognitive processes and mental representations – and the clarification of their role. The result is the cognitive theory of multimedia learning that is represented in the flow chart in Figure 3.2 of this chapter.

The primary addition represented in this chapter is the triarchic model of three demands on cognitive capacity (summarized in Table 3.6) and the

three learning scenarios (summarized in Figure 3.3). These elements seek to link the cognitive theory of multimedia learning to an instructional framework; that is, the goal of these additional elements is to bridge the science of learning (represented in Figure 3.2) with the science of instruction (represented in the three kinds of instructional goals summarized in Table 3.7).

The Present: Progress Report

In the first edition of this handbook (Mayer, 2005), I called for work in (1) fleshing out the details of the mechanisms underlying the five cognitive processes and the five forms of representation, (2) integrating the various theories of multimedia learning, and (3) building a credible research base. In the ensuing decade, we have seen important progress on each of these goals. First, studying the mechanisms of cognitive processing during multimedia learning has been aided by the increasing use of new methodologies, including eye-tracking techniques (e.g., Johnson & Mayer, 2012; Scheiter & van Gog, 2009). Second, the theoretical focus has been strengthened by a focus on the three demands on cognitive capacity (as summarized in Table 3.6) as an organizing and unifying theme. Third, the research base has grown dramatically, as is indicated by the growing number of meta-analyses (Ginns, 2005, 2006; Ginns, Martin, & Marsh, 2013) and by the increasing focus on boundary conditions – that is, pinpointing the conditions under which design principles are more or less likely to apply, including the role of the learner's prior knowledge (see Chapter 24) and the learner's working memory capacity (see Chapter 25).

The Future: Incorporating Motivation and Metacognition

How will the cognitive theory of multimedia learning evolve? A useful next step would be to better incorporate the role of motivation and metacognition in multimedia learning. The rationale for this suggestion is that in addition to being able to engage in appropriate cognitive processing during multimedia learning, successful learners must want to engage in appropriate cognitive processing (i.e., motivation) and know how to manage their cognitive processing (i.e., metacognition).

Motivation to learn (which can be called *academic motivation*) refers to a learner's internal state that initiates and maintains goal-directed behavior (Mayer, 2011). According to this definition, academic motivation is (1) personal (i.e., it occurs within a learner), (2) activating (i.e., it initiates learning behavior), (3) energizing (i.e., it fosters persistence and intensity during learning), and (4) directed (i.e., it is aimed at accomplishing a learning goal). In sum, motivation to learn is reflected in the learner's willingness to exert effort to engage in appropriate cognitive processing during learning (such

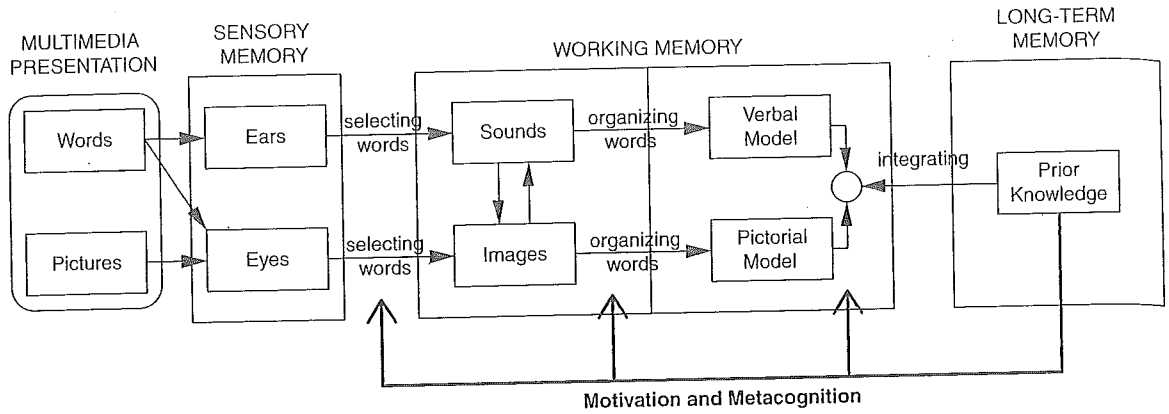


Figure 3.4. *Incorporating motivation and metacognition into a cognitive theory of multimedia learning.*

as the processes of selecting, organizing, and integrating that are needed for meaningful learning).

Metacognition in multimedia learning refers to the learner's awareness and control of cognitive processing during learning (Mayer, 2011). Metacognition plays a crucial role in multimedia learning by guiding the learner's cognitive processing during learning, such as when a learner knows which cognitive activity would be best for a particular learning task and adjusts cognitive activity on the basis of how well it is helping learning. In short, effective multimedia learning includes helping learners become self-regulated learners – that is, learners who take responsibility for managing their cognitive processing during learning.

Although the learner's motivation to learn is part of the definition of generative processing (as summarized in Table 3.6), the overall role of motivation and metacognition is an underdeveloped aspect of the cognitive theory of multimedia learning (Mayer, 2014). Moreno's (2007; Moreno & Mayer, 2007) cognitive affective theory of learning with media seeks to expand multimedia learning theory by more explicitly incorporating the role of motivation and metacognition, highlighted by adding arrows from long-term memory back to the cognitive processing arrows of selecting, organizing, and integrating. Consistent with this approach, Figure 3.4 (adapted from Mayer, 2011) presents a modified version of the cognitive theory of multimedia learning that takes a preliminary step in acknowledging the role of motivation and metacognition in multimedia learning by adding arrows from long-term memory back to the cognitive processing arrows of selecting, organizing, and integrating.

Future research is needed to spell out in greater detail the mechanisms of motivation and metacognition (i.e., how the added arrows work) and to test relevant instructional techniques for promoting academic motivation, such as using emotional design principles to create appealing but relevant graphics (Um, Plass, Hayward, & Homer, 2011). In addition, work is needed

to develop dependent measures for learning outcomes, including the use of delayed tests of retention and transfer; and better measures of the learning process, including measures of cognitive load, motivation, and metacognitive control. Methodological advances including EEG, fMRI, eye-tracking methods, and physiological measures may contribute to these efforts.

Conclusion

In summary, multimedia learning takes place within the learner's information processing system – a system that contains separate channels for visual and verbal processing, a system with serious limitations on the capacity of each channel, and a system that requires appropriate cognitive processing in each channel for active learning to occur. In particular, multimedia learning is a demanding process that requires selecting relevant words and images, organizing them into coherent verbal and pictorial representations, and integrating the verbal and pictorial representations with each other and with relevant prior knowledge. In the process of multimedia learning, material is represented in five forms – as words and pictures in a multimedia presentation, acoustic and iconic representations in sensory memory, sounds and images in working memory, verbal and pictorial models in working memory, and knowledge in long-term memory. During learning, cognitive capacity must be allocated among extraneous, essential, and generative processing, so the goal of instructional design is to develop effective techniques for reducing extraneous processing, managing essential processing, and fostering generative processing.

The theme of this chapter is that multimedia messages should be designed to facilitate multimedia learning processes. Multimedia messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not. This proposition is tested empirically in the chapters of this handbook.

Glossary

Cognitive theory of multimedia learning: A theory of how people learn from words and pictures, based on the idea that people possess separate channels for processing verbal and visual material (dual-channel assumption), each channel can process only a small amount of material at a time (limited-capacity assumption), and meaningful learning involves engaging in appropriate cognitive processing during learning (active processing assumption).

Essential processing: Cognitive processing during learning that is needed to represent the essential presented material in working memory and is caused by the complexity of the material.

Extraneous processing: Cognitive processing during learning that does not serve the instructional objective and is caused by poor instructional design.

Generative processing: Cognitive processing during learning that is aimed at making sense of the essential material in the lesson and is caused by the learner's motivation to exert effort.

Integrating: A cognitive process in which the learner builds connections between visual and verbal representations in working memory and between them and relevant prior knowledge activated from long-term memory.

Long-term memory: A memory store that holds large amounts of knowledge over long periods of time.

Multimedia instructional message: A communication containing words and pictures intended to foster learning.

Multimedia principle: People learn more deeply from words and pictures than from words alone.

Organizing: A cognitive process in which the learner mentally arranges the incoming information in working memory into a coherent cognitive representation.

Selecting: A cognitive process in which the learner pays attention to relevant incoming material and transfers it to working memory for further processing.

Sensory memory: A memory store that holds pictures and printed text impinging on the eyes as exact visual images for a very brief period and that holds spoken words and other sounds impinging on the ears as exact auditory images for a very brief period.

Working memory: A limited-capacity memory store for holding and manipulating sounds and images in active consciousness.

Acknowledgments

I wish to thank Jeroen van Merriënboer, John Sweller, and Wolfgang Scnotz for their many useful comments. Preparation of this chapter was supported by a grant from the Office of Naval Research. The chapter is a revision of chapter 3 of the first edition of *The Cambridge Handbook of Multimedia Learning*.

References

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559.

Baddeley, A. D. (1999). *Human memory*. Boston: Allyn & Bacon.

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Hove: Psychology Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn.* Washington, DC: National Academy Press.

Chambliss, M. J., & Calfee, R. C. (1998). *Textbooks for learning*. Oxford: Blackwell.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*, 149–210.

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology*, *80*, 448–456.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*, 313–332.

Ginns, P. (2006). Integrating information: A meta-analysis of spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*, 511–525.

Ginns, P., Martin, A. J., & Marsh, H. M. (2013). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review*, *25*, 445–472.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.) (2009). *Handbook of metacognition in education*. New York: Routledge.

Johnson, C., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *18*, 178–191.

Kintsch, W. (1998). *Comprehension.* New York: Cambridge University Press.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, *81*, 240–246.

Mayer, R. E. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology*, *84*, 405–412.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*, 1–19.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2002). Multimedia learning. In B. H. Ross (Ed.), *The psychology of learning and motivation*, vol. 41 (pp. 85–139). San Diego, CA: Academic Press.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction*, *12*, 125–141.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*(8), 760–769.

Mayer, R. E. (2009). *Multimedia learning* (2d ed). New York: Cambridge University Press.

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction*, *24*, 171–173.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of the dual-coding hypothesis. *Journal of Educational Psychology*, *83*, 484–490.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, *84*, 444–452.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, *88*, 64–73.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*, 715–726.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*, 187–198.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*, 312–320.

Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology*, *91*, 638–643.

Mayer, R. E., & Sims, V. K., (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*, 389–401.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research & Development*, *43*, 31–43.

Miller, G. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Moreno, R. (2007). Optimising learning from animations by minimizing cognitive load: Cognitive and affective consequences of signaling and segmentation methods. *Applied Cognitive Psychology*, *21*, 765–781.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, *92*, 117–125.

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*, 309–326.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford, England: Oxford University Press.

Paivio, A. (2006). *Mind and its evolution: A dual coding approach*. Mahwah, NJ: Lawrence Erlbaum.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, *90*, 25–36.

Plass, J. L., Moreno, R., & Brunken, R. (Eds.) (2010). *Cognitive load theory*. New York: Cambridge University Press.

Scheiter, K., & van Gog, T. (2009). Using eye tracking in applied research to study and stimulate the processing of information from multi-representational sources. *Applied Cognitive Psychology*, *23*, 1209–1214.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*, 141–156.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell: ACER Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2011). Emotional design in multimedia learning. *Journal of Educational Psychology*, *104*, 485–498.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*, 345–376.