I. Introduction to Multimedia Learning

Instructional messages. However, determining what works in multimedia learning is only half the job. We also want to know how the instructional methods work, that is, we need to understand how our instructional manipulations affect people's cognitive processing during learning. These issues are addressed in the next chapter, which focuses on the science of learning.

SUGGESTED READINGS

Asterisk (*) indicates that part of this chapter is based on this publication.

"How Lightning Storms Develop"

"How Brakes Work"

"How Pumps Work"

"How Plants Grow"

The science of learning is concerned with a theory of how people learn. Theory-grounded practice refers to developing instructional methods that are consistent with how people learn. 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. A cognitive theory of multimedia learning assumes that the human information-processing system includes dual channels for visual/pictorial and auditory/verbal processing, each channel has limited capacity for processing, and active learning entails carrying out appropriate cognitive processing during learning. Five steps in multimedia learning are selecting relevant words from the presented text or narration, selecting relevant images from the presented illustrations, organizing the selected words into a coherent verbal representation, organizing the selected images into a coherent visual representation, and integrating the visual and verbal representations and prior knowledge. Processing of pictures occurs mainly in the visual/pictorial channel; processing of spoken words occurs mainly in the auditory/verbal channel; but processing of printed words takes place initially in the visual/pictorial channel and then moves to the auditory/verbal channel. Three kinds of cognitive load are extraneous cognitive processing, which is cognitive processing that does not serve the instructional goal and is caused by poor instructional design; essential processing, which is cognitive processing that is required to represent the material in working memory and is determined by the complexity of the material; and generative processing, which is deep cognitive processing including organizing and integrating the material. Effective instructional design depends on techniques for reducing extraneous processing, managing essential processing, and fostering generative processing.
Chapter Outline

THE SCIENCE OF LEARNING
What Is the Science of Learning?
What Is Theory-Grounded Practice?
What Is Learning?

THREE ASSUMPTIONS OF A COGNITIVE THEORY OF MULTIMEDIA LEARNING
Dual-Channel Assumption
Limited-Capacity Assumption
Active-Processing Assumption

FIVE STEPS IN A COGNITIVE THEORY OF MULTIMEDIA LEARNING
Selecting Relevant Words
Selecting Relevant Images
Organizing Selected Words
Organizing Selected Images
Integrating Word-Based and Image-Based Representations

EXAMPLES OF HOW THREE KINDS OF PRESENTED MATERIALS ARE PROCESSED IN A COGNITIVE THEORY OF MULTIMEDIA LEARNING
Processing of Pictures
Processing of Spoken Words
Processing of Printed Words

THREE KINDS OF COGNITIVE LOAD IN A COGNITIVE THEORY OF MULTIMEDIA LEARNING
Extraneous Cognitive Processing
Essential Cognitive Processing
Generative Cognitive Processing

CONCLUSION

In line with Stokes's (1997, p. 73) call for "use-inspired basic research" (as described in Chapter 1), this book has two goals - to contribute to practice (i.e., the science of instruction), which is addressed in Chapter 2, and to contribute to theory (i.e., the science of learning), which is addressed in this chapter. Overall, this book is concerned with research on multimedia learning principles that meet two criteria: (a) being theory-grounded – the principles are derived from a cognitive theory of multimedia learning – and (b) being evidence-based – the principles are consistent with empirical research on multimedia learning.

The first criterion – being theory-grounded – is introduced in this chapter, in which I spell out a cognitive theory of multimedia learning. The second criterion – being evidence-based – is introduced in Chapter 2, in which I summarize how we developed an empirical research base for multimedia design principles. Both criteria are infused in Chapters 4–13, in which I describe how well twelve instructional design principles work when tested in multimedia learning environments (i.e., evidence-based) and explain how each principle works within the context of a theory of multimedia learning (i.e., theory-grounded).

In particular, in this chapter I spell out a cognitive theory of multimedia learning – that is, a cognitive theory of how people construct knowledge from words and pictures. First, I explore three fundamental assumptions underlying the theory; second, I examine each of five steps in meaningful multimedia learning based on the theory; third, I give examples of how three kinds of materials are processed; and finally, I distinguish among three kinds of cognitive load in multimedia learning.

THE SCIENCE OF LEARNING

What Is the Science of Learning?

How do people learn? This is the central question in the science of learning. The science of learning is concerned with the creation of a theory of learning based on scientific evidence. The cognitive theory of multimedia learning, which I describe in this chapter, is a research-based theory of learning aimed specifically at explaining learning from words and pictures.

What Is Theory-Grounded Practice?

Principles of multimedia instructional design should be based on an understanding of how people learn from words and pictures. This is the premise underlying theory-grounded practice. An advantage of theory-grounded practice is that instead of rigidly following instructional principles, instructors can have a better understanding of how instructional principles work and the conditions under which instructional principles are most likely to be effective.

What Is Learning?

Learning is a change in knowledge attributable to experience. This definition has three parts: (a) learning is a change in the learner; (b) what is changed is the learner's knowledge; and (c) the cause of the
change is the learner’s experience in a learning environment. Learning is personal, in that it happens within the learner’s cognitive system. The change in knowledge cannot be directly observed but must be inferred from a change in the learner’s behavior—such as performance on a test. The change may involve reorganizing and integrating knowledge rather than simply adding new knowledge. What is learned may involve five kinds of knowledge (Anderson et al., 2001; Mayer & Wittrock, 2006):

- **facts**—knowledge about characteristics of things or events, such as “Sacramento is the capital of California.”
- **concepts**—knowledge of categories, principles, or models, such as knowing what a dog is or how a pulley system works,
- **procedures**—knowledge of specific step-by-step processes, such as how to enter data into a spreadsheet,
- **strategies**—knowledge of general methods for orchestrating one’s knowledge to achieve a goal, such as knowing how to break a problem into subparts, and
- **beliefs**—cognitions about oneself or about how one’s learning works, such as the belief that “I am not good at math.”

In short, learning always involves a change in what the learner knows. In this chapter, I explore the idea that what is learned depends on the learner’s cognitive processing during learning.

**THREE ASSUMPTIONS OF A COGNITIVE THEORY OF MULTIMEDIA LEARNING**

The guiding criterion for this chapter is that the design of multimedia environments should be compatible with how people learn. In short, principles of multimedia design should be sensitive to what we know about how people process information.

What is the role of a theory of learning in multimedia design? 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. Designing multimedia messages is always informed 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, 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. Thus, it does not matter which modality is used to present information—for example, presenting words as sounds or as 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, in presenting many isolated pieces of information, this design is based on a passive-processing assumption—humans act as tape recorders who add 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? Research in cognitive psychology paints a quite different view of how the human mind works (Bransford, Brown, & Cocking, 1999; Lambert & McCombs, 1998; Mayer, 2008a). Thus, the difficulty with this commonsense conception of learning is that it conflicts with what is known about how people learn.

Figure 3.1 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. 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 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 period in an auditory sensory memory (at the bottom). The arrow from pictures to
eyes corresponds to a picture being registered in the eyes; the arrow from words to ears corresponds to spoken text being registered in the ears; and the arrow from words to eyes corresponds to printed text being registered in the eyes.

The central work of multimedia learning takes place in working memory, so let's focus there. Working memory is used for temporarily 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 3.1 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 – of which you are consciously aware – takes place in your working memory. The left side of working memory 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, which I called visual and auditory in Chapter 1. By contrast, the right side of working memory represents the knowledge constructed in working memory – pictorial and verbal mental models and links between them – so it is based on the two representation modes, which I called pictorial and verbal in Chapter 1. The arrow from sound images to visual 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 visual images to sound images represents the mental conversion of a visual image (such as a mental picture of a cat or the printed word “cat”) into a sound image (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 in order 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).

In this chapter, I explore three assumptions underlying a cognitive theory of multimedia learning – dual channels, limited capacity, and active processing. These assumptions – which are derived from the learning sciences – are summarized in Table 3.1. In accordance with the dual-channel assumption, I have divided sensory memory and working memory into two channels – the one across the top deals with visual images and eventually with pictorial representations, whereas the one across the bottom deals with auditory sounds and eventually with verbal representations. In this way I try to compromise between the sensory-modality view – which I use to create two channels on the left side of working memory – and the representation-mode view – which I use to create two channels on the right side of working memory.

In accordance with the limited-capacity assumption, working memory is limited in the amount of knowledge it can process at one time – so that only a few images can be held in the visual channel of working memory, and only a few sounds can be held in the auditory channel of working memory. In accordance with the active-processing assumption, I have added arrows to represent cognitive processes for selecting knowledge to be processed in working memory (i.e., arrows labeled selecting, which move from the presented material to working memory), for organizing the material in working memory into coherent structures (i.e., arrows labeled organizing, which move from one kind of representation in working memory to another), and for integrating the created knowledge with other knowledge, including knowledge brought in from long-term memory (i.e., arrows labeled integrating, which move from long-term memory to working memory and between the visual and auditory representations in working memory). The major cognitive processes required for multimedia learning are represented by the arrows labeled selecting images, selecting sounds, organizing images, organizing sounds, and integrating – which are described in the next section.
I. Introduction to Multimedia Learning

Auditory/Verbal Channel Highlighted

Figure 3.2. The auditory/verbal channel (top frame) and the visual/pictorial channel (bottom frame) in a cognitive theory of multimedia learning.

Dual-Channel Assumption

The dual-channel assumption is that humans possess separate information-processing channels for visually represented material and auditorially represented material. The dual-channel assumption is summarized in Figure 3.2: the top frame shows the auditory/verbal channel highlighted, and the bottom frame shows the visual/pictorial channel highlighted. When information is presented to the eyes (such as illustrations, animations, video, or on-screen text), people begin by processing that information in the visual channel; when information is presented to the ears (such as narration or nonverbal sounds), people 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, 1992, 1999).

What Is Processed in Each Channel?

There are two ways of conceptualizing the differences between the two channels – one based on presentation modes and the other based on sensory modalities. The presentation-mode approach focuses on whether the presented stimulus is verbal (such as spoken or printed words) or nonverbal (such as illustrations, video, animation, or background sounds). According to the presentation-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.

By contrast, the sensory-modality approach focuses on whether learners initially process the presented materials through their eyes (such as for illustrations, 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 auditorially represented material. This conceptualization is most consistent with Baddeley’s (1992, 1999) distinction between the visuo-spatial sketchpad and the articulatory (or phonological) loop.

Whereas the presentation-mode approach focuses on the format of the stimulus-as-presented (i.e., verbal or nonverbal), the sensory-modalities approach focuses on the stimulus-as-represented in working memory (i.e., auditory or visual). The major difference concerning multimedia learning rests in the processing of printed words (e.g., on-screen text) and background sounds. On-screen text is initially processed in the verbal channel in the presentation-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 presentation-mode approach but in the auditory channel in the sensory-mode approach.

For purposes of the cognitive theory of multimedia learning, I have opted for a compromise in which I use the sensory-modalities approach to distinguish between visually presented material (such as pictures, animations, video, and on-screen text) and auditorially 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 needed to clarify the nature of the differences between the two channels.

What Is the Relation Between the Channels?

Although information enters the human information system via one channel, learners may also 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 in 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. Such 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 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 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 the 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 (1992, 1999) theory of working memory and Sweller's (1999, 2005a; Chandler and Sweller, 1991) 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 the person a memory span test (Miller, 1956; Simon, 1980). For example, in a digit span test, I can read a list of digits at the rate of one digit per second (such as 8-7-5-3-9-6-4) and ask you to repeat them back in order. The longest list that you can recite without making an error is your memory span for digits (or digit span). Alternatively, I can show you a series of line drawings of simple objects at the rate of one per second (such as moon-pencil-comb-apple-chair-book-pig) and ask you to repeat them back in order. Again, the longest list you can recite without making an error is your memory span for pictures. 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 (i.e., "eight seven five" pause "three nine" pause "six four"). In this way, the cognitive capacity remains the same – that is, five to seven chunks – but more elements can be remembered within each chunk. Researchers have developed more refined measures of verbal and visual working memory capacity, but continue to show that human processing capacity is severely limited.

**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 (1992) calls the central executive – the system that controls the allocation of cognitive resources – and play a central role in modern theories of intelligence (Steenberg, 1990).

**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, organizing incoming information, 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 tape recorders who file copies of their experiences in memory to be retrieved later.

**What Are the Major Ways in Which 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 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, 1980). 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 of two theories of learning with respect to the nature of the learner, teacher, and instructional methods. Generalization structures can be represented as branching trees and consist of a main idea with subordinate supporting details. An example is an outline of a chapter explaining the major causes of the American Civil War. 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 book. Classification structures can be represented as hierarchies and consist of sets and subsets. An example is a biological classification system for sea animals. These structures are summarized in Table 3.2.

<table>
<thead>
<tr>
<th>Type of Structure</th>
<th>Description</th>
<th>Representation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>Explain a cause-and-effect chain</td>
<td>Flow chart</td>
<td>Explanation of how the human ear works</td>
</tr>
<tr>
<td>Comparison</td>
<td>Compare and contrast two or more elements along several dimensions</td>
<td>Matrix</td>
<td>Comparison of two theories of learning with respect to the nature of the learner, teacher, and instructional methods</td>
</tr>
<tr>
<td>Generalization</td>
<td>Describe main idea and supporting details</td>
<td>Branching tree</td>
<td>Presentation of thesis for the major causes of the American Civil War along with evidence</td>
</tr>
<tr>
<td>Enumeration</td>
<td>Present a list of items</td>
<td>List</td>
<td>List of the names of twelve principles of multimedia design</td>
</tr>
<tr>
<td>Classification</td>
<td>Analyze a domain into sets and subsets</td>
<td>Hierarchy</td>
<td>Description of a biological classification system for sea animals</td>
</tr>
</tbody>
</table>

Understanding a multimedia message often involves constructing one of these kinds of knowledge structures. This assumption suggests two important implications for multimedia design: (a) the presented material should have a coherent structure, and (b) the message should provide guidance to the learner on how to build the structure. If the material lacks a coherent structure – if it is, say 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?

Three processes that are essential for active learning are selecting relevant material, organizing selected material, and integrating selected material with existing knowledge (Mayer, 2005a, 2008a, 2008b; Mayer & Wittrock, 2006; 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 — for example, by using one of the five kinds of structures described earlier. 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, learners must pay attention to certain words and images, arrange them into a cause-and-effect chain, and relate the steps to prior knowledge such as the principle that hot air rises. These processes are summarized in Table 3.3.

In sum, the implicit theory of learning underlying some multimedia messages is that learning is a single-channel, unlimited-capacity, passive-processing activity. Thus, multimedia design is sometimes based on the empty-vessel view of learning described in Chapter 1 — the idea that the learner lacks knowledge so learning involves pouring information into the learner’s empty mind. By 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.

FIVE STEPS IN A COGNITIVE THEORY OF MULTIMEDIA LEARNING

Building on the three assumptions described in the previous section, Figure 3.1 presents a cognitive theory of multimedia learning. For purposes of this book, I define a multimedia environment as one in which material is presented in more than one format — such as in words and pictures. For meaningful learning to occur in a multimedia environment, the learner must engage in five cognitive processes: (1) selecting relevant words for processing in verbal working memory, (2) selecting relevant images for processing in visual

### Table 3.3 Three Processes for Active Learning

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting</td>
<td>Learner pays attention to relevant words and pictures in a multimedia message in order to create a word base and an image base.</td>
<td>In viewing a narrated animation on lightning formation, learner pays attention to words and pictures describing each of the main steps.</td>
</tr>
<tr>
<td>Organizing</td>
<td>Learner builds internal connections among selected words in order to create a coherent verbal model and among pictures in order to create a coherent pictorial model.</td>
<td>Learner organizes the steps into a cause-and-effect chain for the words and for the pictures.</td>
</tr>
<tr>
<td>Integrating</td>
<td>Learner builds external connections between the verbal and pictorial models and with prior knowledge.</td>
<td>Learner makes connections between corresponding steps in the verbal chain and in the pictorial chain and justifies the steps on the basis of his or her knowledge of electricity.</td>
</tr>
</tbody>
</table>

... working memory, (3) organizing selected words into a verbal mental model, (4) organizing selected images into a visual mental model, and (5) integrating verbal and visual representations. 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. More research is needed to clarify how these processes are monitored and coordinated.

Selecting Relevant Words

The first labeled step shown in Figure 3.1 involves a change in knowledge representation from the external presentation of spoken words (such as a computer-generated narration) to a sensory representation of sounds, and then to an internal working memory representation of word sounds. The input for this step is a spoken verbal message — that
is, the spoken words in the multimedia message. The output for this step is a word sound base - 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”). If the words are presented as on-screen text or printed text, however, 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” on the left side of working memory). The need for selecting only part of the presented message arises 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, 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 - for example, “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, and then 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 is selecting relevant images, which 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, so learners must focus on only part of the incoming pictorial material. 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 out 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 may be 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 words and phrases selected from the incoming verbal message - and the output for this step is a verbal model - a coherent (or structured) representation of the selected words or phrases in the learner’s working memory.

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 processes. 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, the learner might 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 pictures selected from the incoming pictorial message – and the output for this step is a pictorial model – a coherent (or structured) representation of the selected images in the learner’s working memory.

This change from images to a 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 image base, but rather must focus on building a simple set of connections. Like the process of organizing words, the process of organizing images is not arbitrary. Rather, it reflects an effort toward building 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 that 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 Image-Based Representations**

Perhaps the most crucial step in multimedia learning involves making connections between word-based and image-based representations. This step involves a change from having two separate representations – a pictorial model and a verbal 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 that is based on connecting the two representations. In addition, the pictorial and verbal models are connected with prior knowledge activated from long-term memory.

I refer to this cognitive process as integrating because it involves building connections between corresponding portions of the pictorial and verbal models as well as knowledge from long-term memory. This process occurs in visual and verbal working memory, and involves the coordination between them. This is a demanding process that requires the efficient use of cognitive capacity. The process reflects the epiteme of sense-making because the learner must focus on the underlying structure of the pictorial and verbal representations. The learner can use prior knowledge 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 visual chain – the blue arrow followed by the red arrow followed by the cloud shape. In addition, prior knowledge can be applied to the transition from the first to the second event by remembering that hot air rises.

Each of the five steps in multimedia learning is likely to occur many times throughout a multimedia presentation. The steps are applied segment by segment – not on the entire 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 visual 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.

In short, multimedia learning takes place in 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 coordinated cognitive processing in each channel in order 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 prior knowledge. The theme of this book is that multimedia messages
I. Introduction to Multimedia Learning

should be designed to facilitate these 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 following ten chapters.

EXAMPLES OF HOW THREE KINDS OF PRESENTED MATERIALS ARE PROCESSED IN A COGNITIVE THEORY OF MULTIMEDIA LEARNING

Let’s take a closer look at how three kinds of presented materials are processed from start to finish according to the model of multimedia learning summarized in Figure 3.1: pictures, spoken words, and printed words. For example, suppose that Albert clicks on an entry for lightning in a multimedia encyclopedia and is presented with a static picture of a lightning storm accompanied by a paragraph of on-screen text about the number of injuries and deaths caused by lightning each year. Similarly, suppose that Barbara clicks on the entry for lightning in another multimedia encyclopedia and is presented with a short animation along with narration describing the steps in lightning formation. In these examples, Albert’s presentation contains static pictures and printed words, whereas Barbara’s presentation contains dynamic pictures and spoken words.

Processing of Pictures

The top frame in Figure 3.3 shows the path for processing of pictures – indicated by thick arrows and darkened boxes. The first event – represented by the pictures box under MULTIMEDIA PRESENTATION on the left side of Figure 3.3 – is the presentation of Albert’s lightning photograph (i.e., a static picture) or Barbara’s lightning animation (i.e., a dynamic picture). The second event – represented by the eyes box under SENSORY MEMORY – is that the pictures impinge on the eyes, resulting in a brief sensory image – that is, for a brief time Albert’s eyes behold the photograph and Barbara’s eyes behold the animation frames. These first two events happen without much effort on the part of the learner, but now the active cognitive processing begins – the processing over which the learner has some conscious control. If Albert pays attention to the fleeting image coming into his eyes (or Barbara attends to the images coming into her eyes), parts of the image will become represented in working memory; this attentional processing corresponds to the arrow labeled selecting images, and the resulting mental representation is labeled images under WORKING MEMORY. Once the visual base is full of image pieces, the next active cognitive processing involves organizing those pieces into a coherent structure – a process indicated by the organizing images arrow. The resulting knowledge representation is a pictorial model – that is, Albert builds an organized visual representation of the main parts of a lightning bolt, or Barbara builds an organized set of images representing the cause-and-effect steps in lightning formation. Finally, active cognitive processing is required to connect the new representation with other knowledge – a process indicated by the integrating arrow. For example, Albert may use his prior knowledge about electricity to help him include moving positive and negative charges in his mental representation, or Barbara may use her prior knowledge of electricity to help explain why the negative and positive charges are attracted to one another. In addition, if the learners have also produced a verbal
mental model, they may try to connect it to the pictorial model—for example, by looking for how a phrase in the text corresponds to a part of the image. This processing results in an integrated learning outcome, as indicated by the circle under working memory.

Processing of Spoken Words

The middle frame in Figure 3.3 shows the path for processing of spoken words—indicated by thick arrows and darkened boxes. When the computer produces spoken narration (as indicated by the words box under multimedia presentation), the sounds are picked up by Barbara's ears (as indicated by the ears box under sensory memory). For example, when the computer says, "The negatively charged particles fall to the bottom of the cloud, and most of the positively charged particles rise to the top," these words are picked up by Barbara's ears and held temporarily in her auditory sensory memory. Next, active cognitive processing can take place. If she pays attention to the sounds coming into her ears (as indicated by the arrow labeled selecting words), some of the incoming sounds will be selected for inclusion in Barbara's word sound base (indicated by the sounds box under working memory). For example, the resulting collection of words in working memory might include: "positive top, negative bottom." The words in the word base are disorganized fragments, so the next step—indicated by the organizing words arrow—is to build them into a coherent mental structure—indicated by the verbal model box. In this process, the words change from being represented based on sound to being represented based on word meaning; the result could be a cause-and-effect chain for the steps in lightning formation. Lastly, Barbara may use her prior knowledge to help explain the transition from one step to another and may connect words with pictures—such as connecting "positive top, negative bottom" with an image of positive particles in the top of a cloud and negative charges in the bottom. This process is labeled integrating, and the resulting integrated learning outcome is indicated by the circle under working memory.

Processing of Printed Words

So far, cognitive processing of pictures takes place mainly in the visual/pictorial channel (shown in the bottom half of Figure 3.1), whereas the cognitive processing of spoken words takes place mainly in the auditory/verbal channel (shown in the top half of Figure 3.1). However, the arrow from images to sounds indicates that the learner (such as Barbara) can mentally create sounds corresponding to the visual image—such as mentally saying "wind" when she sees wavy arrows in the animation. Similarly, the arrow from sounds to images indicates that the learner (such as Barbara) can mentally create images corresponding to the word sound base—such as visualizing a plus sign when the narration says "positively charged particle."

The presentation of printed text in multimedia messages seems to create an information-processing challenge for the dual-channel system portrayed in Figure 3.1. For example, consider the case of Alan, who must read text and view an illustration. The words are presented visually, so they must initially be processed through the eyes—as indicated by the arrow from words to eyes. Then, Alan may attend to some of the incoming words (as indicated by the selecting images arrow) and bring them into working memory as part of the visual image base. Then, by mentally pronouncing the images of the printed words Alan can get the words into the word sound base—as indicated by the arrow from images to sounds. Once the words are represented in the auditory/verbal channel they are processed like spoken words, as described earlier. This path is presented in the bottom frame of Figure 3.3. As you can see, when verbal material must enter through the visual channel, the words must take a complex route through the system, and must also compete for attention with the illustration that Alan is also processing via the visual channel. The consequences of this problem are addressed in more detail in Chapter 11 on the modality principle.

THREE KINDS OF COGNITIVE LOAD IN A COGNITIVE THEORY OF MULTIMEDIA LEARNING

A central tenet of the cognitive theory of multimedia learning (Mayer, 2005a, 2008b; Mayer & Moreno, 2003) and cognitive load theory (Chandler & Sweller, 1991; Clark, Nguyen, & Sweller, 2006; Sweller, 1999, 2005a) from which it was derived is that learners can engage in three kinds of cognitive processing during learning, each of which draws on the learner's available cognitive capacity. Table 3.4 summarizes the three kinds of cognitive load, which DeLeeuw and Mayer (2008) refer to as the triarchic model of cognitive load.

Extraneous Cognitive Processing

As indicated in the first row of Table 3.4, extraneous cognitive processing (which Sweller [1999] calls extraneous cognitive load) refers to
cognitive processing during learning that does not serve the instructional goal and that is caused by confusing instructional design. For example, if captions are printed at the bottom of the screen and an animation is presented above, the learner will have to visually scan back and forth between words at the bottom of the screen and the corresponding portion of the animation. This visual scanning is a form of extraneous processing because it wastes precious cognitive capacity due to poor design. The poor design can be corrected by placing words next to the portion of the graphic they describe. If extraneous processing consumes all of the learner’s available cognitive capacity, then the learner is not able to engage in cognitive processes for learning such as selecting, organizing, and integrating. The result is no learning, which is reflected in poor retention and poor transfer performance.

**Essential Cognitive Processing**

As indicated in the second row of Table 3.4, essential cognitive processing (which Sweller [1999] calls *intrinsic cognitive load*) is cognitive processing during learning that serves to represent the essential material in working memory and that is determined by the inherent complexity of the material. For example, for a learner who is unfamiliar with the material, the lightning lesson shown in Figure 2.2 is so complex—consisting of many steps and underlying processes—that it could overload the learner’s cognitive capacity. One way to help learners process complex material is to provide pre-training in the names and characteristics of the key elements. Essential processing corresponds to the selecting arrows in Figure 3.1, which indicate that the learner is building a representation of the material in working memory. If learners engage mainly in essential cognitive processing during learning, the result will be rote learning, as reflected in good retention and poor transfer performance.

**Generative Cognitive Processing**

As indicated in the third line of Table 3.4, generative cognitive processing (which Sweller [1999] calls *germane cognitive load*) is cognitive processing during learning that is aimed at making sense of the essential material and that can be attributed to the learner’s level of motivation. Generative processing corresponds to the organizing and integrating arrows in Figure 3.1, which indicate deeper processing. Generative processing may be primed by creating an engaging learning environment in which the narrator uses a conversational style and polite wording. If learners are able to engage in essential and generative processing, they are more likely to construct a meaningful learning outcome that enables both good retention and good transfer performance.

According to this triarchic model of cognitive load, a major challenge of instructional design is that cognitive capacity is limited, so there is only a limited capacity for extraneous, essential, and generative processing. You can see that each of the three kinds of demands on cognitive capacity leads to a different problem for instructional design: problems attributable to confusing design of the physical layout of the material, problems attributable to the inherent complexity of the material, and problems attributable to unmotivating communication style (Mayer, 2008b).

The three kinds of problems respectively require three kinds of instructional design solutions: reduce extraneous cognitive processing, manage essential cognitive processing, and foster generative cognitive processing. The first kind of problem occurs when confusing instructional design encourages the learner to engage in extraneous cognitive processing, thereby limiting the amount of cognitive capacity available for essential and generative processing. To combat this problem, instructional designers should design lessons that reduce extraneous processing in learners. The next section of this book (Chapters 4-8) reviews five techniques for reducing extraneous cognitive processing—the coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles.

The second kind of problem occurs when essential processing consumes all or most of the learner’s cognitive capacity, leaving insufficient capacity available for generative processing. To address
this problem, we need lessons that manage essential processing. The third section of the book (Chapters 9-11) reviews three techniques for managing essential processing – the segmenting, pre-training, and modality principles.

Finally, even if capacity is available for generative processing, we have a problem if learners are not motivated to use that capacity. To address this problem, we need to design lessons that foster generative processing. The fourth section of this book (Chapters 12-13) examines two techniques aimed at fostering generative processing – the multimedia principle and the personalization principle. It also examines the role of the narrator’s voice and image on learning (i.e., the voice and image principles).

CONCLUSION

The theme of this chapter is that the design of multimedia messages should be consistent with a research-based theory of how people learn. In this chapter, I presented a cognitive theory of multimedia learning based on three well-established ideas in cognitive science – what I call the dual-channel, limited-capacity, and active-learning processing assumptions. I showed how multimedia learning occurs when the learner engages in five kinds of processing – selecting words, selecting images, organizing words, organizing images, and integrating. I gave examples of how pictures, spoken words, and printed words are processed according to the cognitive theory of multimedia learning. Finally, I showed how instructional designers need to reduce extraneous processing, manage essential processing, and foster generative processing. In the remainder of this book, I use the cognitive theory of multimedia learning to suggest design principles that my colleagues and I have tested.

SUGGESTED READINGS

Asterisk (*) indicates that a portion of this chapter was based on this publication.