

Attention, Working Memory, and Long-Term Memory in Multimedia Learning: An Integrated Perspective Based on Process Models of Working Memory

Judith Schweppe · Ralf Rummel

Published online: 25 October 2013

© Springer Science+Business Media New York 2013

Abstract Cognitive models of multimedia learning such as the Cognitive Theory of Multimedia Learning (Mayer 2009) or the Cognitive Load Theory (Sweller 1999) are based on different cognitive models of working memory (e.g., Baddeley 1986) and long-term memory. The current paper describes a working memory model that has recently gained popularity in basic research: the embedded-processes model (Cowan 1999). The embedded-processes model argues that working memory is not a separate cognitive system but is the activated part of long-term memory. A subset of activated long-term memory is assumed to be particularly highlighted and is termed the “focus of attention.” This model thus integrates working memory, long-term memory, and (voluntary and involuntary) attention, and referring to it within multimedia models provides the opportunity to model all these learning-relevant cognitive processes and systems in a unitary way. We make suggestions for incorporating this model into theories of multimedia learning. On this basis, one cannot only reinterpret crucial phenomena in multimedia learning that are attributed to working memory (the split-attention effect, the modality effect, the coherence effect, the signaling effect, the redundancy effect, and the expertise reversal effect) but also derive new predictions.

Keywords Multimedia learning · Working memory · Process models · Attention

Cognitive theories of multimedia learning such as the Cognitive Load Theory (CLT; Sweller 1999; Sweller et al. 2011) or the Cognitive Theory of Multimedia Learning (CTML; Mayer 2009) are grounded in empirical findings and theoretical principles regarding cognitive architecture and cognitive processes. As their cognitive foundation, these models thus rely on the assumptions of basic research, referring to perception, sensory memory, attention, working memory, and long-term memory. From all these research areas, mainly those assumptions are included that are regarded as widely accepted knowledge in the respective field (e.g., that there is a limited-capacity working memory system with a pictorial and a verbal subsystem and an unlimited long-term memory). When specific models are referred to, these are preferentially the most influential ones. A particular focus is set on working memory, since it is assumed that “the central work of multimedia learning takes place in working memory” (Mayer 2005, p. 38): temporally maintaining, manipulating, and integrating verbal and pictorial information. Due to this emphasis,

J. Schweppe (✉) · R. Rummel

Department of Psychology, University of Erfurt, P. O. Box 900 221, 99105 Erfurt, Germany
e-mail: judith.schweppe@uni-erfurt.de

a central recommendation for the design of multimedia learning materials is to take into account the limited capacity of working memory. For a long time, the prevailing working memory model in both basic and applied research contexts has been the multicomponent model (Baddeley and Hitch 1974; Baddeley 1986), which conceives of working memory as a *cognitive system* separate from long-term memory that is responsible for briefly maintaining and manipulating information. This view is also the one incorporated in multimedia models.

Interestingly, in basic research on working memory, the multiple systems view is being increasingly questioned and models are put forward that rather conceive of working memory as *attentional processes* operating on *long-term memory* (e.g., Barrouillet et al. 2004; Cowan 1984, 1995, 1999; Oberauer and Kliegl 2006; Ruchkin et al. 2003; for similar ideas, see already Anderson 1983; Hebb 1949; James 1890). Among these models, the *embedded-processes model* of memory by Cowan (1995, 1999) is the most prominent and also most general one. In this review, we advocate the application of such process models in general and of the embedded-processes model in particular to multimedia learning. In our view, an incorporation of such a model has benefits for the theories of multimedia learning. One advantage is the stronger focus on long-term memory, which has been shown to influence performance in working memory tasks to a much stronger degree than a focus on working memory as a separate system suggests. Another, more important, advantage is the opportunity to explain and predict learning-relevant mechanisms based on attentional processes, working memory and long-term memory within *one* memory model. So far, these components of the cognitive architecture subserving learning are included in models of multimedia learning, but in a rather isolated way. We argue that referring to a more unitary basic memory model brings conceptual clarity and allows for deriving new hypotheses, in particular regarding boundary conditions of well-known effects in multimedia learning. Moreover, this model provides a more sophisticated concept of attention by distinguishing task- and stimulus-driven allocation of attention, as will be explained below in the description of the embedded-processes model. Based on this distinction, inconsistencies in the literature can be better explained and new predictions can be made. The inclusion of an alternative view of memory in general and working memory in particular thus functions as an “update” of the cognitive foundation of models of multimedia learning, which serves the purpose of aligning these with more recent theoretical assumptions and empirical findings in basic research and also makes new predictions regarding instructional practice.

We will start by summarizing the role of attention, working memory, and long-term memory in the two most influential theories of multimedia learning, the CTML (e.g., Mayer 2009) and the CLT (e.g., Sweller et al. 2011). After a sketch of Baddeley’s (1986) working memory model and problems of this type of model, we will then describe process models of working memory and, in particular, discuss their potential implications for multimedia learning. In this context, we will describe how this different concept of working memory can be incorporated in the CTML and in the CLT. Furthermore, we will provide alternative explanations of important empirical findings regarding the design of multimedia learning materials that have been attributed to working memory as well as new hypotheses—in particular regarding attention—that can be derived from a *process* approach to working memory.

Attention, Working Memory, and Long-Term Memory in Theories of Multimedia Learning

Both the CTML and the CLT argue that learning is constrained by the characteristics of human cognitive architecture: Working memory is said to be a store that allows for “temporally holding

and manipulating knowledge in active consciousness” (Mayer 2005, p. 37). This implies that all working memory processes require attention and equates the allocation of working memory resources and the allocation of attention. While working memory is limited in its capacity as well as temporally, long-term memory has no such limitations.

Specifically, the CTML (e.g. Mayer 2009) assumes verbal and pictorial information to be processed as follows (see also Fig. 1): Incoming information is briefly stored in sensory memory, depending on the input modality in a visual or in an auditory sensory store. Relevant visual and auditory information is then selected and transferred to modality-specific subsystems of working memory, where it can be maintained and processed. Each of these modality-specific working memory stores is limited in capacity. In the course of processing, knowledge in terms of a pictorial and a verbal model of the incoming information is constructed in working memory and then integrated with each other and with prior knowledge (represented in long-term memory). Once integration has occurred, the goal of learning has been achieved. Since both organizing and integrating information draw on the limited capacity of working memory, it is essential for learning to avoid working memory overload during processing. A way to avoid this is to distribute memory load in such a way that none of the subsystems is overloaded. In contrast, long-term memory is relevant only in terms of providing background knowledge that comes into play in the final stage of situation model building.

CLT also relies on the assumption of multiple memory stores with a capacity-limited working memory and an extensive long-term memory and focuses on working memory and cognitive load therein. Cognitive load represents the working memory (or attentional) resources that learning makes demand on (e.g. Sweller and Chandler 1994). Three types of cognitive load are distinguished in the CLT, which result from the learner’s interaction with the learning material: intrinsic cognitive load, which depends on the complexity of the content to-be-learned in relation to the learner’s prior knowledge, in particular on the number of elements that must be simultaneously processed in working memory to enable comprehension; extraneous cognitive load, which is caused by the cognitive demands imposed by instructional design, for instance, when visual search processes must be executed; and finally, germane cognitive load, which results from engaging in activities that foster integration with prior knowledge and the construction of schemata, for instance, when learners are required to self-explain their processing steps in problem solving (e.g., Renkl 2002). As these types of cognitive load are assumed to be additive, learning performance should be best when extraneous load is reduced, thus allowing for free cognitive capacity to be invested as germane load (Sweller et al. 2011). According to CLT, long-term memory contributes to the learning process such that cognitive load can be reduced by the contribution of schemata. Since multiple elements can be organized in one schema, the opportunity to draw on pre-existing organized knowledge reduces the demands on working

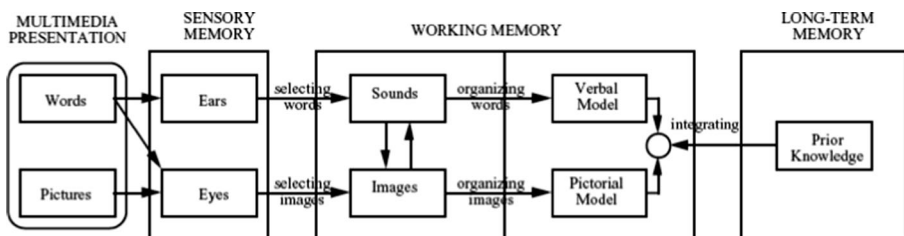


Fig. 1 The cognitive theory of multimedia learning. The figure is taken from Mayer (2009, p. 61)

memory. It is however emphasized that the influence of schemata is usually limited in instruction because of the need to deal with novel information.

It is evident that both the CTML and the CLT are based on the widely established multiple memory stores approach, according to which short-term/working memory and long-term memory constitute separate cognitive systems (e.g. Atkinson and Shiffrin 1968). In line with this, Baddeley and Hitch's (1974) multicomponent model describes a working memory system, which is separated from long-term memory. It further distinguishes a central executive, an attentional control system, and two slave systems, one dedicated to the storage and processing of verbal information, the phonological loop, and the other dedicated to the storage and processing of (nonverbal) visuo-spatial information, the visuo-spatial sketchpad. As a consequence of overwhelming evidence that long-term memory strongly influences performance in working memory tasks (see also below), Baddeley (2000) modified his model by introducing the episodic buffer as a new component, which acts as an interface between long-term memory and the (other) subsystems of working memory. Still, this updated version of the model maintains the structural separation of working memory and long-term memory.

Although the memory assumptions in the CTML and the CLT allow for explaining a large range of findings, some problems arise from the current concept of working memory and the relation to long-term memory assumed therein. Firstly, the allocation of working memory resources to certain processes is equated with the allocation of attention. Yet, it is unclear how this assumption of central attentional limits is related to the assumption of modality-specific capacity limits. Specifying the role of attention in the process of learning with multimedia materials is thus a desideratum for further research. In addition, it is widely acknowledged that prior knowledge is an important moderator of learning, but its exact role still remains to be specified. In particular, in the CTML, prior knowledge exerts an influence only at the final stage of integration, whereas the influence of long-term knowledge on prior stages of processing is not modeled. Neither do we know precisely how the integration of prior knowledge from long-term memory and novel information stored in working memory takes place. Furthermore, it remains an open but important question how information is transferred from working memory to long-term memory. Even though it is claimed that the processes central to learning take place in working memory, it is evident that the newly acquired information needs to be established as a long-term memory trace. In particular, in the CTML, this final process is not specified either—a description of the processing steps typically ends with building connections between novel verbal and pictorial information and prior knowledge (e.g., Mayer 2009). The current conception of attention, working memory, and long-term memory in models of multimedia learning therefore leaves several questions unanswered that are central to learning.

The process models that we discuss here emphasize that working memory constitutes a subset of the representations of long-term memory, some of which are highlighted by (voluntary and involuntary) attentional mechanisms and thus are particularly focused on. We argue that, if applied to theories of multimedia learning, the rejection of the idea of separate systems as well as the focus on attentional mechanisms provides the opportunity to address the just-described problems and to unify the assumptions that relate to attention, working memory, and long-term memory.

As mentioned above, the most influential of these models is the embedded-processes model of working memory (Cowan 1995, 1999), which is also broad in its coverage and is conceived as an integrative, unitary model of working memory, long-term memory, and attention. Given the unsolved questions brought up above, this integrative approach may also be useful for modeling the cognitive processes underlying multimedia learning. We will therefore focus on this model, though many assumptions apply to other process models (e.g., Barrouillet and Camos 2012; Barrouillet et al. 2004; Majerus et al. 2004; Oberauer and Kliegl 2006; Ruchkin et al. 2003; Zimmer 2008) as well.

Nelson Cowan's (1995, 1999) Embedded-Processes Model of Memory

According to the embedded-processes model of working memory (Cowan 1995, 1999), only sensory memory and long-term memory are to be distinguished in terms of memory systems. Working memory is thought of as “cognitive processes that retain information in an unusually accessible state” (Cowan 1999, p. 62). What contributes to working memory is information in long-term memory in general, the currently activated part of long-term memory and the subset of activated memory that is in the *focus of attention*. How easily accessible information is, thus depends on its state of activation and *not* on its being stored in a specific system: “The information in the focus is the most readily accessible information in working memory. Information that is activated, but not to the point of conscious awareness, also can be retrieved reliably [...] Finally, an item that is not even in an activated state still can be thought of as in working memory [...] if there are cues in working memory that point to the item and raise the likelihood that it could be retrieved if necessary” (Cowan 1999, p. 65f). Thus, even items that are not activated but that are closely associated to currently activated representations can be accessed more easily than other representations in long-term memory, which also implies that only a subset of working memory is consciously experienced. The main difference to systems models is thus that there is no separate, additional storage system for information “in” working memory, but rather increased activation of the respective long-term memory information.

During encoding, a stimulus activates multiple features in long-term memory. If the stimulus is attended to or recruits attention, more features are activated than if not, including semantic ones. This richer encoding results in a more robust memory representation. Given the influence of attention, a crucial question is what determines whether or not information is in the focus of attention. In line with theories of attention (e.g., Pashler 1988; Posner 1980), Cowan (1999) distinguishes two modes of attention allocation: changing, novel, or salient stimuli automatically recruit attention via an attentional orienting mechanism; in addition, attention is allocated voluntarily via executive control so that those representations are highlighted that are most useful for the task. Via the central executive attention cannot only be directed inward to long-term memory representations but also outward to stimuli in the environment. This outward direction of attention takes place during visual search or while trying to follow a certain voice in a dichotic listening task (or in a multispeaker environment). Thus, the focus of attention is jointly determined by voluntary and involuntary mechanisms (see Fig. 2).

Forgetting in working memory is assumed to be due to a time limit in activated long-term memory (via decay and interference) and to a capacity limit of the focus of attention. The amount of activation is not limited, however. For the time limit, the relative and not the absolute amount of time is assumed to be essential because, during pauses, reactivation processes can occur that counteract forgetting. In line with this assumption, performance is better when the ratio between the pauses between successive stimuli and the retention interval is larger. In addition, interference is particularly strong—though not restricted to cases—when representations are similar on any possible level. This is called similarity-based interference. With respect to the attentional capacity limit, Cowan (1999, 2001) suggests that four (plus/minus 1) unconnected items or chunks can be kept in the focus of attention at one time, irrespective of the modality in which the information has originally been processed (for an overview reporting converging evidence from several domains for this new “magical number” as an update to Miller’s (1956) magical number of seven plus/minus two chunks, see Cowan 2001). It is evident that short-term memory span is usually higher than four items, but this is because tasks such as digit or word span not only tap the focus of attention

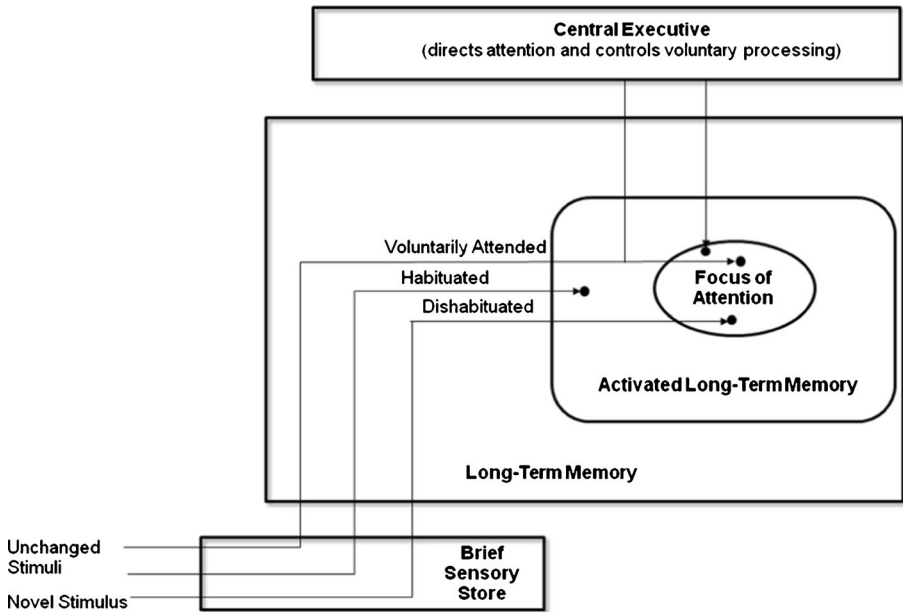


Fig. 2 Processing of unchanged and novel/salient stimuli in the embedded-processes model. Unchanged stimuli do not automatically recruit attention, but can be voluntarily attended to via the Central Executive; novel or salient stimuli automatically recruit attention. Figure adapted from Cowan (1999)

but also reflect the contribution of activated long-term memory. It is thus only the focus of attention that is limited in capacity but not working memory in general.

In order to counteract forgetting, some form of maintenance mechanism is necessary. In Cowan's (1999) model, maintenance is mainly accomplished via central executive processes. As a functional analogue to verbal rehearsal, which is the main maintenance mechanism in Baddeley's (1986) model, items in the focus of attention are reactivated via a mental search through a set of items. It is further assumed that not all features activated during encoding will be reactivated (and their activation thus prolonged). Depending on the task demands, attention will be focused so that the most useful representations are selected for reactivation. The critical difference to articulatory rehearsal (and also to a visuo-spatial maintenance mechanism) in Baddeley's model is that these processes apply to various types of representation alike and that they tap the central executive and therefore demand cognitive resources. With respect to encoding and passive storage as well as with respect to active maintenance, the embedded processes model thus makes a less sharp distinction between verbal and visuo-spatial sources of information and regards them simply as "two varieties of memory activation" (Cowan 1999, p. 71).

The assumption that working memory and long-term memory are structurally intertwined also makes the transition from working memory to long-term memory a critical aspect of the theory. It is assumed that, when "novel" information is encoded in long-term memory, the single elements are not new but rather the combination of elements previously stored in long-term memory as singletons. An example for this is the combination of the single sounds and syllables that make up a new word or the combination of two propositions that constitute the message of a sentence. In line with the idea that declarative memories are encoded only with the presence of attention (Cowan 1995), it is assumed that a combination of elements that are currently in the focus of attention may be stored as a new long-term memory trace.

Empirical Evidence in Favor of Process Models

In contrast to models that assume structurally separate stores for short-term memory/working memory on the one hand and long-term memory on the other hand (e.g., Atkinson and Shiffrin 1968; Baddeley and Hitch 1974), the approach just described makes the crucial prediction that the same representations are involved in the processing and in the retention of information. Furthermore, it follows that mechanisms supporting maintenance apply to verbal and visual stimuli alike. Thus, both verbal and visual working memory should be supported by domain-general task-related and stimulus-related attentional mechanisms. Both predictions are supported by behavioral and neuropsychological evidence and respective evidence is briefly reported in the following.

In the verbal domain, there are numerous studies demonstrating that performance in verbal working memory tasks is influenced by the organization of verbal long-term memory and the mechanisms of language processing (e.g., Acheson and MacDonald 2009; Dell et al. 1997; Majerus et al. 2004; Martin and Saffran 1997; Martin et al. 1999; Schwegge and Rummel 2007). Several findings indicate that verbal stimuli are indeed encoded at various levels and that a phonological code alone cannot account for verbal working memory, as would be the case if the phonological loop was the basic storage device underlying verbal working memory performance. Not only phonological representations contribute to simple verbal span tasks but also lexical and semantic ones: For instance, lists of words with high frequency of occurrence are recalled better than lists of low-frequency words (Watkins and Watkins 1977) and recall for lists of concrete words is better than recall of abstract words (Walker and Hulme 1999). The frequency effect is interpreted in terms of frequent words being more easily accessible in the mental lexicon as compared to infrequent words. The concreteness effect suggests that the richer semantic (long-term memory) representation of concrete over abstract words (e.g., De Groot 1989; Paivio 1971) also supports their short-term maintenance. When sentences or texts instead of unconnected words are to be remembered, conceptual and syntactic representations (such as the message or the structure of a sentence, respectively) additionally come into play and are even more important than phonology. In a seminal study, Sachs (1967) demonstrated that lexical–phonological changes in a to-be-recognized sentence that preserved the sentence’s meaning were detected only immediately after sentence presentation. Syntactic and semantic changes were reliably detected after a considerably longer delay (see also Lombardi and Potter 1992; Potter and Lombardi 1990, 1998; Rummel et al. 2003). Furthermore, the fact that far more words within a sentence than unrelated words in a list can be retained (sentence superiority effect; Brener 1940) argues for the assumption that knowledge structures in long-term memory that are related to currently presented stimuli (such as knowledge about what word order makes a grammatically acceptable sentence) also contribute to performance in working memory tasks even though they are only indirectly activated by the stimuli. The sentence superiority effect is also observed in the presence of an attention-demanding secondary task, which makes it impossible to attribute the effect to an effortful integration of long-term memory and working memory representations (Baddeley et al. 2009). The close relationship between verbal long-term memory and verbal working memory demonstrated in these studies thus supports the assumption that working memory is embedded in long-term memory. This is augmented by findings that the very same cortical areas that participate in perception and production of verbal stimuli are also involved in their storage and that there is an increase in neural synchrony between attentional control systems (in prefrontal cortex) and those areas (for an overview, see Ruchkin et al. 2003).

Analogous specifications exist for visuo-spatial working memory, assuming that it consists of executive processes operating on visuo-spatial long-term memory representations (Postle 2006; Zimmer 2008). For instance, Chinese letters are often used as stimuli with Western participants because they are complex visual stimuli for which little long-term knowledge can be assumed. In contrast, Chinese participants ought to possess long-term representations of these letter symbols and of the radicals that constitute them. If visual working memory is activated long-term memory for visual information, they should have a higher memory span for these stimuli even for pseudo-letters consisting of existing radicals in new combinations (that can thus not be named). Zimmer and Fu (2008) demonstrated that Chinese participants indeed outperformed German participants in retaining Chinese pseudo-letter symbols. However, with visual stimuli that neither group was familiar with (nonsense figures) their performance was equal. In neuropsychological studies, there is overwhelming evidence that “the same posterior neural structures that are involved in perception and contribute to long-term memory” (Zimmer 2008, p. 1374) are involved in visuo-spatial working memory. Analogous to verbal working memory, there is additional activation in prefrontal structures, which serve to direct attention (Courtney 2004).

With respect to attentional control, the idea is that short-term memory for all kinds of stimuli is supported by the same attentional mechanisms and that voluntary and involuntary attentional control can be distinguished. And, indeed, the “same attention mechanisms that have been shown to shape the neural networks of visual STM [short-term memory] also shape those of verbal STM [short-term memory]” (Majerus et al. 2012). A dorsal attention network supports task-related attention and, in terms of Cowan’s (1999) model, is responsible for the voluntary maintenance of information in the focus of attention. A ventral attention network supports stimulus-related attention and is activated by salient and unexpected stimuli irrespective of the task—corresponding to the attentional orienting system in the embedded processes model. Neuroimaging studies have also demonstrated that an increase in the memory load, which affords higher attentional maintenance, leads (1) to an increase in activation in the dorsal attention network interpreted as task-related, voluntary attention and (2) to a decrease in activation in the ventral attention network interpreted as stimulus-related attention, that is, the recruitment of attention for novel or salient stimuli. These patterns of activation and deactivation occur both in verbal and in visual short-term memory tasks (Majerus et al. 2012; Todd et al. 2005; Todd and Marois 2004). Load-dependent activation of the dorsal attention network reflects an attentional pointer function, “connecting with distinct representational substrates, as a function of the type of information to be pointed to” (Majerus et al. 2012). The decrease in activation of the ventral attention network as a function of memory load reflects reduced reactivity to irrelevant stimuli so that the chance for irrelevant items being attended to and thus interfering with the relevant stimuli is minimized. Both attentional functions—strengthening activation of relevant stimuli and weakening the chance of activation of irrelevant stimuli—are at play for visual as well as verbal stimuli and become increasingly important with increasing load.

These findings from basic research (both behavioral and neural) demonstrate that there is good reason to favor process models that emphasize the structural embeddedness of working memory within long-term memory and the important role of attention for short-term maintenance. When compared to separate system models, these models can equally account for the low interference between visual and verbal working memory that was pivotal for the multicomponent model by referring to the differences in long-term knowledge structures. They can better account for the finding that other features than phonological ones are important for verbal working memory and that the same—cognitive and neural—representations are involved in the processing and maintenance of verbal as well as visual stimuli. In addition,

the finding that task-related and stimulus-related attentional mechanisms that are involved in retention can be distinguished and that they apply to verbal and visual stimuli alike is as predicted by the embedded-processes model (Cowan 1999).

In the following sections, we will outline how this alternative concept of working memory can be incorporated in both the CTML and the CLT. Finally, we will turn to the question how findings in multimedia research that are attributed to attention and working memory and their limitations can be reinterpreted in the light of a process view of memory and what further predictions follow from the incorporation of such a working memory model.

What Does This Imply for Theories of Multimedia Learning?

In the following, we will outline how this different concept of working memory can be incorporated in the CTML and the CLT and what changes this implies, both for the theories and for instructional predictions.

The CTML as a Process Model

There are three main differences between the concept of working memory as it is applied in the CTML and the concept of working memory in the embedded-processes model: (1) the CTML structurally separates working memory and long-term memory; the embedded-processes model assumes a single system but distinguishes different states of activation; (2) in the CTML, it is assumed that all processes in working memory require attention, and thus, the allocation of working memory resources is not distinguished from the allocation of attention; in the embedded-processes model, it is assumed that only a subset of the information in working memory is attended to; and (3) only the embedded-processes model distinguishes between involuntary, stimulus-driven and voluntary, task-driven allocation of attention.

Adopting a process perspective on working memory within the CTML thus affects not only the working memory component but requires a more general incorporation of the embedded-processes model in the CTML because attention, working memory, and long-term memory are affected. The most central change is a change of what “is” working memory: instead of being represented as a separate system between sensory memory and long-term memory, working memory needs to be represented “in” the cognitive processes—selecting words/images, organizing words/images, and integrating in terms of the CTML (see Fig. 1 above), and activation and attention in terms of the embedded-processes model. “Selecting” is equivalent to activating representations in long-term memory as well as to the additional recruitment of attention when “these activated features are sufficient” (Cowan 1999, p. 65). “Organizing” and “integrating” need to be conceptualized as *one* process because the encoding of novel information always takes place in long-term memory and therefore necessarily includes the integration with pre-existing knowledge.

An entirely new distinction that should be made is that between (involuntary) stimulus-driven attention and (voluntary) task-driven attention. All stimuli that are being processed activate representations on different levels in long-term memory but attended stimuli activate more features. The activated representations can be attentionally focused either because they recruit attention due to salience or novelty of the stimulus or because attention is directed to them via executive control. Therefore, a central executive, which is also a core part of Baddeley’s (1986) working memory model, needs to be incorporated in the CTML. Via this

central executive, attention is directed either inward, that is, to certain long-term memory representations, or outward, that is, to certain stimuli in the environment. Attention serves to encode the stimuli more strongly or to maintain their activation above threshold and thus to counteract forgetting. More specifically, in the CTML, the outward direction of attention is represented in the so far unlabeled arrows/processes between the multimedia presentation and sensory memory; the inward direction and the stimulus-driven recruitment of attention are subsumed under “selecting words/images” (see Fig. 1).

Theoretical Implications for CLT: Specifying Cognitive Load

We suggest that conceiving of working memory as attentional processes (and activation) operating on long-term memory also provides the opportunity to determine extraneous and intrinsic cognitive load more specifically and in relation to the capacity of the focus of attention. Within CLT, intrinsic cognitive load is characterized as the cognitive demands resulting from the complexity of the task in relation to the prior knowledge of the learner, and extraneous cognitive load is regarded as a consequence of the cognitive activities that are not essential for learning but that are imposed by the design. Recently, Kalyuga (2011) argued that only intrinsic and extraneous load should be distinguished and germane load is redundant with intrinsic load. According to Kalyuga (2011, p. 3), intrinsic load “is associated with processing the essential interacting elements of information and their relations that define the corresponding schematic structures,” and thus, it is essential to learning just as it is defined for germane load. On the basis of the embedded-processes model, a reinterpretation of intrinsic and extraneous cognitive load in relation to attentional processes is possible.

Intrinsic Cognitive Load According to Sweller (1999), element interactivity is assumed to be the decisive factor that determines the intrinsic complexity of materials, and the consequences of instructional manipulations are said to depend on the degree of element interactivity. “Low element interactivity materials allow individual elements to be learned with minimal reference to other elements and so impose a low working memory load. [...] High element interactivity material consists of elements that heavily interact and so cannot be learned in isolation” (Sweller 2010, p. 124). It is therefore desirable to be able to specify or even quantify different levels of element interactivity, in particular specify when element interactivity is *too* high. The embedded-processes account provides first steps for such a specification: Since encoding of a new declarative long-term memory trace requires the to-be-combined elements to be simultaneously attended to, the capacity limit of the focus of attention of four (plus/minus one) chunks (Cowan 2001) should be critical here. This suggests that no more than four unrelated elements (irrespective of whether they belong to the multimedia message or to prior knowledge) should need to be combined in order to enable optimal encoding in long-term memory. Whenever this number is exceeded, materials are too complex. An obstacle for an exact, a priori quantification of element interactivity is however the problem to determine the (individual) size of a chunk, which also depends on the individual’s prior knowledge. Furthermore, prior knowledge may not only affect intrinsic cognitive load via chunk size but also via a better organized long-term memory in which both activated and attended representations contribute to understanding and learning with multimedia materials. In spite of these unknowns, relating element interactivity to the capacity of the focus of attention is a promising first step in this direction.

Extraneous Cognitive Load If learning in terms of encoding new long-term memory traces depends on to-be-related elements being simultaneously attended to, extraneous cognitive load

can be regarded as “everything that precludes relevant representations from being attended to.” A critical new distinction that cannot be made on the basis of CLT’s current assumptions concerning human cognitive architecture is whether this is (1) because of irrelevant stimuli attracting involuntary attention or (2) because of an inefficient voluntary allocation of attention via executive processes. Considering the first of these possibilities: if irrelevant information attracts the learner’s attention involuntarily, fewer relevant representations can be in the focus of attention. As suggested by CLT, this should result in poorer learning when element interactivity is high. Considering the second possibility: if the design makes it hard for the learner to realize which information to focus on, the voluntary allocation of attention is impaired. Likewise, if attention-demanding but learning-irrelevant cognitive processes need to be executed (such as visual search), this interferes with the direction of attention to relevant representations. Instructional implications of this distinction will be discussed in the following section.

How Can the Embedded-Processes Model as a Part of Multimedia Theories Explain What We Already Know About (Working) Memory in Multimedia Learning and What Further Predictions Does the Application of this Theory Make?

According to Popper (1935), the application of a new theory requires that it is also able to account for findings that could be accounted for in the previous, established one. Therefore, it is essential to explain known findings in multimedia learning that have been attributed to working memory on the basis of a process view. In addition, a new theory needs to provide new hypotheses that could not have been derived on the basis of the established theory. As argued in the beginning, we are positive that process models can be useful for research on multimedia learning in that they provide an integrated framework for attention, working memory, and long-term memory within which existing phenomena can be accounted for and on the basis of which new hypotheses can be deduced. In the following, we will turn to these implications in detail. We have identified central findings that are attributed to working memory and its limited capacity within the CTML and/or CLT and will address each of them in turn. These are the split-attention effect (Ayres and Sweller 2005; Kalyuga et al. 2003; Moreno and Mayer 1999), the modality effect (Mayer and Moreno 1998; Mousavi et al. 1995) and its boundary conditions, the seductive details (or coherence) effect (Mayer et al. 2001), the signaling effect (Mautone and Mayer 2001), the redundancy effect (Mayer et al. 2001), and the expertise reversal effect (Kalyuga et al. 2003). At the end of this section, we additionally sketch potential implications for how individual differences in working memory capacity may affect the benefits of instructional guidance. We exemplarily focus on capacity-dependent effects of seductive details and split attention. Partially, the new explanations resemble the old ones, but the main difference is that they can be derived from a single, integrated cognitive architecture. In other cases, the reinterpretation of the effects refers to different processes than previous explanations of the findings. New hypotheses that derive specifically from the embedded-processes model of working memory are described in the course of the reinterpretations and as a first step mainly concern boundary conditions and moderating factors.

The Split-Attention Effect

Learning performance increases, the closer verbal and pictorial information are presented in time and space. Thus, pictures or animations in combination with spoken text or with spatially integrated written text lead to better performance than graphics in combination

with spatially separated written text. In addition, simultaneous presentation of spoken text and graphics is superior to their sequential presentation. These findings are referred to as the split-attention effect, and the respective design recommendations are called the temporal and spatial contiguity principles (Kalyuga et al. 1999; Moreno and Mayer 1999; for an overview, see Ayres and Sweller 2005). The main idea behind the split-attention effect is that the opportunity to attend to two sources simultaneously supports integration. In terms of Cowan's model, to-be-integrated elements need to be simultaneously in the focus of attention if they are to be encoded as a new long-term memory trace. This should be the case when they are processed and attended to in close succession (i.e., when a picture and a text can be processed almost simultaneously). Yet, when the to-be-integrated pieces of information enter the cognitive system time-lagged (i.e., with sequential presentation or when the two stimuli need to be processed sequentially because of sensory limitations), additional attentional rehearsal (via the central executive) is necessary. In addition, attention-demanding processes like visual search that need to be executed for spatially separated texts and graphics interfere with the maintenance of previously encountered information.

Another advantage of materials in which spoken text is presented simultaneously with graphics is that the voluntary allocation of attention to the picture or animation can be guided by those parts of the spoken text that refer to the pictorial information. However, whether the learner can benefit from the attention-guiding function of the spoken text should strongly depend on how easily verbal-pictorial associations can be detected: If it takes time to identify the parts of the picture just referred to, the transient nature of spoken presentation (Leahy and Sweller 2011; Wong et al. 2012) even creates the time-lag it was supposed to avoid. The need for visual search thus not only demands attention but also creates non-contiguity. A prediction that follows from this is that split-attention effects in comparing spoken and written texts accompanying graphics should not be found with strong demands for visual search, a pattern that was indeed observed by Jeung et al. (1997).

Stimulus-driven allocation of attention as assumed by Cowan (1999) may be another stumbling block for the benefits of simultaneous processing of spoken and pictorial information: If salient parts of a picture or particularly of an animation that are currently not being referred to in the text catch attention (e.g., by motion), integration of verbal and pictorial information should be hindered, and thus, no split-attention effect should occur.

The Modality Effect and its Boundary Conditions

The finding that learning with auditory text and pictures or animations leads to better results than learning with visual texts and pictures or animations is called the modality effect (e.g., Mayer and Moreno 1998; Moreno and Mayer 1999; Mousavi et al. 1995). The prevailing explanation for this effect is that by presenting texts and graphics in two sensory modalities (auditory and visual) rather than one (texts and pictures visual), more capacity in working memory is available because two different subsystems of working memory are addressed initially.

This explanation has been criticized because it does not accord with Baddeley's (1986) codality—specific rather than modality-specific conception of the verbal (auditory and visual) and (nonverbal) visuo-spatial subsystems of working memory (e.g., Rummer et al. 2010; Schüler et al. 2011, 2012; Tabbers 2002) and because there is no evidence that picture processing interferes with reading (Gyselinck et al. 2002; Rummer et al. 2010, 2011). It is also not in accord with the process model focused on here, in which it is only the supramodal focus of attention that is limited in *capacity*. It is thus not the number of auditory and visual pieces of

information that determines the limits of capacity but the overall number of pieces, more precisely of chunks, of information. Consequently, one cannot push the limits of capacity by addressing more than one modality. Yet, within Cowan's (1999) model, the capacity limit of the focus of attention is not the only cause of forgetting: The activation of representations in long-term memory that are not in the attentional focus is subject to decay and interference. Representations are assumed to interfere with each other more strongly when they are more similar. Additional modality-specific interference between written text and pictures as compared to spoken text and pictures might thus be an explanation for the modality effect in Cowan's (1999) model. There is, however, no evidence that the long-term memory representations that are activated in the course of picture processing resemble those activated in the course of reading. Rather, written text is as soon as possible processed in a way specific to language, that is, in terms, of orthographic, phonological, and semantic representations (for an overview, see Fürstenberg et al. 2013). On the basis of this evidence, it is thus not plausible to explain the modality effect in terms of modality-specific interference between activated representations in long-term memory, since the processing of written speech requires the activation (and the maintenance of activation) of language-specific rather than of modality-specific representations.

A widely recognized alternative or additional explanation of the modality effect when it occurs with simultaneous presentation of texts and pictures does not directly refer to working memory but to learners' sensory limitations and their consequences for processing in working memory: The eyes cannot focus on written text and a picture at the same time, but the eyes can focus on a picture while the ears process incoming speech (e.g., Rummer et al. 2011). When texts and pictures are presented simultaneously, the modality effect can thus be interpreted as a special case of the split-attention effect without the need to make any further assumptions. The sensory limitations are particularly harmful when the pace of presentation is fast and not under control of the learner because processing of one source can then occur at the cost of processing the other. The sensory-based explanation can thus also account for the fact that the modality effect does often not occur or even reverse when the pacing of materials is learner-determined (Tabbers 2002; Tabbers et al. 2004).

In addition, acoustic-sensory representations are available for processing longer than visual-sensory ones (Penney 1989). Such differences between acoustic and visual sensory codes are also assumed by Cowan (1999), and they can account for modality effects with brief verbal materials: The availability of an auditory representation of the last words heard should be particularly helpful when the verbal sequences are short, since then the last words constitute a larger proportion of the text that is to be integrated with a picture (see also Rummer et al. 2011; Wong et al. 2012). This assumption can also account for a discrepancy in the literature. With sequential presentation of texts and pictures (i.e., in a condition in which split attention cannot cause an auditory advantage), Mousavi et al. (1995) and Moreno and Mayer (1999, Exp. 2) who presented combinations of pictures and brief texts still observed a modality effect. In contrast, Baggett and Ehrenfeucht (1983) and Tiene (2000) who presented longer texts in combination with pictures did not find any auditory advantage. The text length that is critical here may either be the length of the entire verbal material or of that portion that needs to be integrated with pictorial information prior to the onset of the next part of the text.¹

¹ The finding that no modality effect is found with sequential presentation of longer texts and pictures is also in line with the assumption that long auditory texts are disadvantageous because the learner cannot skip back to previous parts of the text (transient information effect; Leahy and Sweller 2011; Wong et al. 2012).

Within the embedded-processes view, the modality effect might thus be explained in terms of modality-specific interference between long-term memory representations. However, there is no evidence that written texts and pictures activate similar representations in the course of processing. An alternative explanation for the modality effect that also accounts for its boundary conditions has already been proposed in terms of perceptual restrictions.

The Seductive Details Effect

The fact that task-irrelevant but interesting information that is added to a multimedia presentation interferes with learning is also attributed to working memory and its limited capacity. It is referred to as the seductive details effect, and the recommendation to eliminate such nonessential material is called the coherence principle (e.g., Mayer et al. 2001; Mayer 2005). Here, the embedded-processes model provides a straightforward explanation: When irrelevant information captures the learner's attention, it will enter the focus of attention and thus occupy some of the capacity that is required for learning.

As described above, information can enter the focus of attention in two ways: either because it is voluntarily attended to by the central executive or because it automatically attracts attention based on its novelty or salience. The voluntary direction of attention to irrelevant materials should occur when learners are not able to distinguish between relevant and irrelevant information and cannot direct attention efficiently or when their processing is guided by goals other than the acquisition of knowledge. Both ways in which voluntary attention allocation is hindered have direct implications for whether or not the inclusion of seductive details affects learning: the former implies that the seductive details effect is particularly strong for learners with low prior knowledge who are less able to distinguish relevant and irrelevant information than learners with higher prior knowledge; the latter implies that it is particularly strong for learners with low motivation who may be willing to direct their attention to interesting but irrelevant rather than to (potentially less interesting but) learning-relevant information. In this case, the "learner's" goals that drive attention are not necessarily learning-related. Furthermore, an involuntary attraction of attention by seductive details should occur when the irrelevant material is either more *salient* or more *novel* than the relevant information. This attention-grabbing aspect is also inherent to the term "*seductive* details." The negative effect of seductive details should thus be the stronger, the more novel and salient the irrelevant information is compared to the relevant information. Under such circumstances it can hardly be avoided that irrelevant thoughts enter the focus of attention.

The different ways by which items may enter the focus of attention thus provide unique hypotheses for the conditions in which seductive details should substantially hinder learning and conditions in which their damage should be limited. Since these hypotheses are new, they have not yet been tested purposefully. There is, however, some evidence in the literature that supports the relevance of these boundary conditions: Mayer et al. (2008) demonstrated that highly interesting details decreased learning more strongly than low-interest details. Situational interest has been related to factors that are supposed to attract attention involuntarily (e.g., topic shifts, Hidi 1990; suspense, Jose and Brewer 1984; unexpectedness of main events, Iran-Nejad 1987; and emotionally charged or provocative information, Kintsch 1980; Schank 1979; see also Schraw and Lehman 2001). Nonetheless, the interestingness manipulation of Mayer et al. (2008) cannot be entirely equated with salience so that these findings are only a first indicator. Schraw and Lehman (2001) argue that one reason for inconsistent findings regarding the influence of seductive details is that they differ (unsystematically) in their potential to distract

the learner and that the relation between the relevant and the seductive parts of the materials may be important here. The current hypotheses may provide the basis for a systematic investigation of what makes external details seductive or nonseductive.

The idea that participants with low prior knowledge should be particularly susceptible to seductive details, which is based on our current assumptions, has also been voiced by Lowe (2004) and Magner et al. (2013) indeed find such a pattern with decorative illustrations as seductive details in geometry problems. These findings are first indicators that the hypotheses derived from a process view of working memory in multimedia learning are valid, but further research that directly tests these hypotheses and that manipulates the relative salience of irrelevant and relevant parts of the materials is needed.

The Signaling Effect

According to the coherence principle, learners can be helped to identify relevant information by leaving out irrelevant information. A way to direct attention to particularly relevant information—when elimination is not possible or not sensible—is to present visual or verbal cues as signals. The respective recommendation is called “signaling principle” (e.g., Mautone and Mayer 2001).

A central assumption of the embedded-processes model is that stimuli that are attended to (or that recruit attention) activate more of their features in long-term memory and are thus processed more deeply than unattended stimuli. Cues that guide the learner’s attention can thus ensure that the important parts of a text or a picture are encoded more strongly. In addition, the formation of declarative long-term memory traces is assumed to require attention. It is thus essential for long-term learning that the to-be-learned information is processed in the focus of attention. Cowan (1999, p. 78) assumes that “the orienting mechanism and its habituation allow an effortless mode of attention,” while the executive control of attention is effortful. This implies that paying attention is less effortful for a physically salient stimulus than if attention is directed away from the stimulus. Apart from being able to account for signaling effects in general, a model such as Cowan’s (1999) thus makes a further prediction, namely that signaling cues should be particularly helpful when important information would probably escape the learner’s attention without a cue—either because the learner cannot recognize its importance without the help of a cue or because it would otherwise be overlooked. Consequently, important but small changes from one picture to another or within an animation should be made salient because they would not trigger an orienting response without a cue. Complementarily, signaling should have only a little or no beneficial effect if the cued items are salient even without the cue (be it a visual one or a verbal emphasis) and would thus probably be attended to anyway. Furthermore, the assumption that habituation prevents the involuntary allocation of attention also has direct instructional implications for the use of cues in that they would lose their efficiency if used too often and should thus be used selectively to remain effective.

These hypotheses echo assumptions made in the model of animation processing by Lowe and Boucheix (2008); see also Boucheix and Lowe (2010) model of animation processing, in which discrepancies between perceptual salience and thematic relevance are named as a critical hindering factor for learning with animations. Furthermore, de Koning et al. (2009) concluded from their review of research that cues in animations are effective only when the animations are complex or ill-structured (see also Dodd and Antonenko 2012; Jeung et al. 1997). This supports our assumption that cues are only beneficial when the cued elements would otherwise have been overlooked. De Koning et al. (2009, p. 131) further conclude that cues are ineffective

when they “are overridden by attention-catching elements in an animation,” a finding that perfectly fits the assumptions described here. The authors conclude their review by saying that “the lack of a comprehensive theory that encompasses all perceptual and cognitive aspects in learning from cued and uncued animations does not allow researchers to fully explain the effects of visual cueing. Future research should be aimed at integrating the predominantly cognitive accounts with current theories of visual cognition as a more appropriate framework for investigating how cueing might work in processing animations” (de Koning et al. 2009, p. 135). In our view, the incorporation of a working memory model that unifies attention, working memory, and long-term memory, such as Cowan’s (1999) embedded-processes model into theories of multimedia learning is a promising step in this direction and has the potential to account for at least some of these inconsistencies.

The Redundancy Effect

According to the redundancy principle, the presentation of the same content twice in different forms should be avoided because the matching and integration of the redundant pieces of information increases working memory load. Within the CTML, the redundancy effect is restricted to the simultaneous presentation of spoken and written text (Mayer et al. 2001), whereas CLT speaks of redundancy effects more generally whenever two (external or internal) sources provide comparable information, for instance, when information is given that the learner knew before (Chandler and Sweller 1991). The explanation of the redundancy effect implies that two redundant sources of information are particularly harmful when learners do not know that they could ignore one of the sources without losing any information or when they are not able to ignore it. According to Sweller (2005), the limitations of working memory only apply when novel information is to be processed. This distinction between novel information in working memory and existing knowledge in long-term memory is modified from the perspective of a process model of memory. According to this approach, the formation of a new long-term memory trace is the encoding of a novel combination of previously singular elements in long-term memory and depends on these elements being simultaneously in the focus of attention. The problems that are caused by redundant pieces of information that are not recognized as such or that cannot be ignored can easily be accounted for within such an approach: Given the postulated capacity limit of the focus of attention of approximately four chunks, the ability to integrate the relevant information and to encode new long-term memory traces that consist of a combination of several chunks is lowered whenever additional stimuli recruit—and thus enter the focus of—attention. Since multiple features are activated, this can even be the case not only when exactly the same meaning is conveyed in different forms but also when the two sources provide the information in different ways, for instance in paraphrases. If, however, the learner is able to chunk the different versions of the same information such as his prior knowledge and the incoming information, the capacity limit of the focus of attention should no longer be problematic because the information provided by two sources could henceforth be processed as one chunk. A potential implication of this assumption is then that the redundancy effect for redundant but not entirely overlapping information is reduced whenever enough time is available to chunk the similar pieces of information.

A reversal of the redundancy effect has been reported for a case in which not the entire spoken information is repeated in written form, but only the key points and close to the corresponding parts of the graphics (Yue et al. 2013; Mayer and Johnson 2008). In this case, the “redundant” information may well have functioned as a *signal* that highlights the

important portions of the verbal message as well as of the picture. The salient presentation of essential words would then increase the chance that the key aspects of the multimedia message are not only activated in memory but are also attended to.

The Expertise Reversal Effect

An important moderator of many influencing factors in multimedia learning is the prior knowledge of the learners. The finding that most effects in multimedia learning are restricted to learners with low prior knowledge and either reverse or make no difference with high prior knowledge learners is referred to as “expertise reversal effect.” In the CLT, it is explained in terms of the redundancy effect: instructional guidance is redundant with and thus interferes with pre-existing schemata in experts’ long-term memory (Kalyuga et al. 2003). On the basis of the embedded-processes model, this explanation would need to be modified. The activation of prior knowledge is an obligatory component of the encoding of new information. The more a learner knows prior to instruction, the easier it is to build chunks. As far as instructional guidance serves to reduce the number of elements in the focus of attention, the advantages should be less pronounced with higher prior knowledge learners, since their existing knowledge already allows them to build larger chunks. In addition, they can benefit more strongly from activated long-term memory and from associations in long-term memory because their knowledge is better organized. The improvements through instructional guidance may thus not be necessary for high prior knowledge learners, which can explain the finding of no differences between conditions for such participants though not necessarily the reversal.

Apart from the working memory-based explanation, there are however other explanations for expertise reversal effects. For instance, learners with high prior knowledge may be bored with easier, unchallenging materials and thus process materials with more guidance less deeply than more difficult and challenging ones (e.g., Schnitz et al. 2009). In other cases, the contents of the multimedia presentation that aim at helping novices may literally conflict with what experts know about the matter. Under these circumstances, the high prior knowledge learners may be annoyed by the additional contents and engage in (attention-demanding) processes other than knowledge acquisition. Such thoughts are reported by the expert participants of Oksa et al. (2010) who read a Shakespearean play in the presence of explanatory notes. Our preliminary conclusion is that the expertise reversal effect is not a unitary phenomenon but goes back to different mechanisms, the relative contribution of which varies as a function of the nature of the instructional manipulation, the learners’ overall motivation, and probably also the precise level of expertise.

Individual Differences in Working Memory Capacity and Their Implications for Multimedia Learning

The focus of attention has also been related to individual differences in working memory capacity, as measured by complex span tasks such as operation span (Turner and Engle 1989) or reading span (Daneman and Carpenter 1980). An influential account in this context is the controlled attention theory of working memory (e.g., Engle and Kane 2004), which assumes that “individuals with greater working memory capacity (WMC) are better able to control or focus their attention than individuals with lesser WMC” (Colflesh and Conway 2007, p. 699). Working memory capacity is thus not measured in terms of the capacities of

subsystems but in terms of the ability to suppress distracting, irrelevant information and to activate and maintain relevant information in the face of distraction. Which of these is more important is still under debate (e.g., Shipstead and Broadway 2013). It has been demonstrated that high capacity participants are better able to suppress to-be-ignored stimuli (as one's name presented in a to-be-ignored channel, selective attention; Conway et al. 2001) and that they are better able to attend to stimuli in two channels (as noticing one's name in one ear, while also attending to the information presented to the other ear, divided attention; Colflesh and Conway 2007).

With regard to multimedia learning, the relation between working memory capacity and the ability to control the focus of attention has interesting implications for individual difference effects. Focusing on relevant information and not being distracted by seductive irrelevant information in a multimedia learning environment should be harder for learners with low compared to high working memory capacity. Such a pattern has indeed been observed by Sanchez and Wiley (2006). They found a seductive details effect only for low working memory capacity learners. In addition, these participants focused on irrelevant pictures more often than learners with high working memory capacity. The interpretation of differences in working memory capacity in terms of differences in the ability to control the focus of attention also suggests individual differences in the split-attention effect. The poorer performance of individuals with low working memory capacity in divided attention tasks (Colflesh and Conway 2007) implies that these learners should be less able to benefit from the fact that spoken text and graphics can be *perceived* simultaneously because their ability to *attend* to them simultaneously is limited. Yet, to our knowledge, working memory capacity has not yet been tested as a moderating factor of the split-attention effect.

Apart from the seductive details effect and the split-attention effect, other multimedia effects may be influenced by working memory capacity. Differences in the ability to control the focus of attention and to maintain task goals have, for instance, affected the benefit from visual cues and from verbal instruction (Skuballa et al. 2012). In our view, relating working memory capacity to the control of the focus of attention is a promising starting point for further research into how individual differences in working memory capacity affect multimedia learning.

Conclusions

The previous paragraphs have demonstrated that an incorporation of a process view of working memory in the CTML and in CLT accounts for important phenomena that are attributed to working memory. In addition, this different concept of working memory as attentional processes operating on long-term memory representations provides novel hypotheses for factors that moderate the split-attention effect, the seductive details effect, and the signaling effect and may yield further hypotheses in the future. In our view, the application of process models to the modeling of cognitive processes in multimedia learning also has important implications for future directions in research on multimedia learning. In particular, since the embedded-processes model unifies working memory, long-term memory, as well as attention, it provides the opportunity to model mechanisms within one theory that previously needed to be treated separately.

One core assumption of process models is that working memory is not conceived as a cognitive system separate from long-term memory but as different states of activation. Thus, long-term memory is incorporated by definition. The fact that these models regard working memory as embedded in long-term memory suggests a stronger focus on long-term memory and its organization in general. Long-term memory is already in the focus when it comes to

the role of schemata and to the importance of prior knowledge within CLT (Sweller 1999). Yet, when what happens in working memory depends on the structure of long-term memory to such a strong degree, causes of interference as well as mechanisms that support the encoding of long-term knowledge and its consolidation should also be focused on.

Another central assumption is that the focus of attention, which is critical for the establishment of new long-term memory traces, is accessed in two ways: Attention is directed voluntarily to (what the learner thinks to be) task-relevant and useful information via the central executive, and attention is recruited involuntarily to salient or novel stimuli via an attentional orienting system. The distinction of stimulus- and task-driven attentional processes has direct implications for the design of multimedia materials: The most important design criterion should be to ensure that the most relevant pieces of information are attentionally focused on and thus processed deeply. This implies that it should be made as probable as possible that relevant information is focused on and as improbable as possible that irrelevant information is focused on.

Attention is thus a more sophisticated concept than in the CTML and the CLT (as well as in the pre-2000 versions of the multicomponent model). The emphasis on central executive processes is not new to this approach but is also an essential component of Baddeley's (1986, 2000) model. However, this component has not been paid much attention to, either in working memory research or in its application to multimedia learning (for a similar argument, see Schüler et al. 2011). The visuo-spatial and verbal subsystems of working memory have been focused on at the expense of the central executive. The fact that both types of working memory models emphasize the central executive makes the incorporation of this component into multimedia models even more important and the mechanisms driving attention an essential part of the learning process.

References

- Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*, 50–68.
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295.
- Atkinson, R. G., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control process. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York, NY: Academic Press.
- Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp 135–146). New York: Cambridge University Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.
- Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*, 438–456.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.
- Baggett, P., & Ehrenfeucht, A. (1983). Encoding and retaining information in the visuals and verbals of an educational movie. *Educational Communication and Technology Journal*, *31*, 23–32.
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource-sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.
- Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science*, *21*, 413–419.
- Boucheix, J.-M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, *20*, 123–135.
- Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, *26*, 467–482.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332.

- Colflesh, G. J. H., & Conway, A. R. A. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review*, *14*, 699–703.
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331–335.
- Courtney, S. M. (2004). Attention and cognitive control as emergent properties of information representation in working memory. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 501–516.
- Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*, 341–370.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford Psychology Series, no. 26. Oxford: Oxford University Press.
- Cowan, N. (1999). An embedded-processes model of working memory. In: A. Miyake, & P. Shah (Eds.). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.
- De Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 824–845.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, *21*, 113–140.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in normal and aphasic speakers. *Psychological Review*, *104*, 801–838.
- Dodd, B. J., & Antonenko, P. D. (2012). Use of signaling to integrate desktop virtual reality and online learning management systems. *Computers & Education*, *59*, 1099–1108.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). New York: Elsevier.
- Fürstenberg, A., Rummer, R., & Schweppe, J. (2013). Does visuo-spatial working memory generally contribute to immediate serial letter recall? Does visuo-spatial working memory generally contribute to immediate serial letter recall? *Memory*, *21*, 722–731.
- Gyselinck, V., Cornoldi, C., Dubois, V., De Beni, R., & Ehrlich, M. (2002). Visuospatial memory and phonological loop in learning from multimedia. *Applied Cognitive Psychology*, *16*, 665–685.
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, *60*, 549–572.
- Iran-Nejad, A. (1987). Cognitive and affective causes of interest and liking. *Journal of Educational Psychology*, *79*, 120–130.
- James, W. (1890). *Principles of psychology*. New York: Holt.
- Jeung, H. J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, *17*, 329–343.
- Jose, P. E., & Brewer, W. F. (1984). Development of story liking: Character identification, suspense, and outcome resolution. *Developmental Psychology*, *20*, 911–924.
- Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*, 1–19.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23–31.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*, 351–371.
- Kintsch, W. (1980). Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics*, *9*, 87–89.
- Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, *25*, 943–951.
- Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*, 713–733.
- Lowe, R. K. (2004). Interrogation of a dynamic visualization during learning. *Learning and Instruction*, *14*, 257–274.
- Lowe, R. K., & Boucheix, J.-M. (2008). Learning from animated diagrams: how are mental models built? In G. Stapleton, J. Howse, & J. Lee (Eds.), *Theory and applications of diagrams* (pp. 266–281). Berlin: Springer.

- Magner, U. I. E., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2013). Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learning and Instruction*. doi:10.1016/j.learninstruc.2012.07.002.
- Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language*, *51*, 297–306.
- Majerus, S., Attout, L., D'Argembeau, A., Degueldre, C., Fias, W., Maquet, P., et al. (2012). Attention supports verbal short-term memory via competition between dorsal and ventral attention networks. *Cerebral Cortex*, *22*, 1086–1097.
- Martin, R. C., Lesch, M. F., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*, 2–39.
- Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, *14*, 641–682.
- Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, *93*, 377–389.
- Mayer, R. E. (2005). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 183–200). Cambridge: Cambridge University Press.
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge: Cambridge University Press.
- Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied*, *14*, 329–339.
- 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., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology*, *100*, 380–386.
- 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.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
- Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*, 358–368.
- Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, *87*, 319–334.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*, 601–626.
- Oksa, A., Kalyuga, S., & Chandler, P. (2010). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science*, *38*, 217–236.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369–378.
- Penney, C. G. (1989). Modality effects and the structure of short term verbal memory. *Memory & Cognition*, *17*, 398–422.
- Popper, K. R. (1935). *Logik der Forschung*. Wien: J. Springer. [Reprinted as *The logic of scientific discovery*. London: Hutchinson, 1959.]
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*, 23–38.
- Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*, 633–654.
- Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*, 265–282.
- Renkl, A. (2002). Learning from worked-out examples: Instructional explanations supplement self-explanations. *Learning and Instruction*, *12*, 529–556.
- Ruchkin, D., Grafman, J., Cameron, K., & Berndt, R. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences*, *26*, 709–728.
- Rummer, R., Engelkamp, J., & Konieczny, L. (2003). The subordination effect: Evidence from self-paced reading and recall. *European Journal of Cognitive Psychology*, *15*, 539–566.
- Rummer, R., Schweppe, J., Fürstenberg, A., Scheiter, K., & Zindler, A. (2011). The perceptual basis of the modality effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *17*, 159–173.

- Rummer, R., Schweppe, J., Fürstenberg, A., Seufert, T., & Brünken, R. (2010). What causes the modality effect in multimedia learning? Testing a specification of the modality assumption. *Applied Cognitive Psychology, 24*, 164–176.
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*, 437–442.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34*, 344–355.
- Schank, R. C. (1979). Interestingness: Controlling influences. *Artificial Intelligence, 12*, 273–297.
- Schnotz, W., Fries, S., & Horz, H. (2009). Motivational aspects of cognitive load theory. In M. Wosnitza, S. A. Karabenick, A. Efklides, & P. Nenniger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 69–96). New York: Hogrefe & Huber.
- Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review, 13*, 23–52.
- Schüler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review, 23*, 389–411.
- Schüler, A., Scheiter, K., Rummer, R., & Gerjets, P. (2012). Explaining the modality effect in multimedia learning: Is it due to a lack of temporal contiguity with written text and pictures? *Learning and Instruction, 22*, 92–102.
- Schweppe, J., & Rummer, R. (2007). Shared representations in language processing and verbal short-term memory: The case of grammatical gender. *Journal of Memory and Language, 56*, 336–356.
- Shipstead, Z., & Broadway, J. M. (2013). Individual differences in working memory capacity and the stroop effect: Do high spans block the words? *Learning and Individual Differences, 26*, 191–195.
- Skuballa, I., Schwonke, R., & Renkl, A. (2012). Learning from narrated animations with different support procedures: Working memory capacity matters. *Applied Cognitive Psychology, 26*, 840–847.
- Sweller, J. (1999). *Instruction design in technical areas*. Camberwell: ACER.
- Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 159–167). Cambridge: Cambridge University Press.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–233.
- Tabbers, H. K. (2002). *The modality of text in multimedia instructions: Refining the design guidelines*. Doctoral dissertation, Open University of The Netherlands, Heerlen.
- Tabbers, H., Martens, R., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*, 71–81.
- Tiene, D. (2000). Sensory load and information load: Examining the effects of timing on multisensory processing. *International Journal of Instructional Media, 27*, 183–199.
- Todd, J. J., Fougine, D., & Marois, R. (2005). Visual short-term memory load suppresses temporo-parietal junction activity and induces inattention blindness. *Psychological Science, 16*, 965–972.
- Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*, 751–754.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.
- Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1256–1271.
- Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 712–718.
- Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction, 22*, 449–457.
- Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology, 105*, 266–277.
- Zimmer, H. D. (2008). Visual and spatial working memory: From boxes to networks. *Neuroscience and Biobehavioral Reviews, 32*, 1373–1395.
- Zimmer, H. D., & Fu, X. (2008). *Working memory capacity and culture-based expertise*. Paper presented at the XXIV International Congress of Psychology, Berlin.