Cognitive Load Theory: Implications for Affective Computing

Slava Kalyuga
School of Education, University of New South Wales
Sydney, NSW 2052 AUSTRALIA
s.kalyuga@unsw.edu.au

Abstract

Cognitive load theory investigates instructional consequences of processing limitations of human cognitive system. Excessive cognitive load inevitably influences our affective states by causing frustration that may discourage further learning activities, while low-load tasks may also be disengaging and boring. This paper reviews basic assumptions of cognitive load theory and their consequences for optimizing the design of information presentations, as well as general implications for affective computing.

Introduction

It is rather obvious from common learning experiences that cognitive load would influence our affective states. Excessively complex and cognitively taxing tasks could frustrate learners and discourage further learning ("too much headache"), while very simple and non-challenging cognitive activities may also disengage learners ("too boring"). In many situations, the observed affective states result from specific cognitive conditions and could be used as indicators of such conditions while modifying learning environments. It has been also demonstrated that emotional states (e.g., negative mood or anxiety) directly influence cognitive task performance and the operation of working memory, while less evidence exists about the effect of the emotional content of the processed information (e.g., Kensinger & Corkin, 2003).

Specific factors influencing cognitive load characteristics of learning and instruction have been investigated for several decades within a framework of cognitive load theory (for a recent comprehensive overview, see Sweller, Ayres, & Kalyuga, 2011). Unfortunately, their consequences for learner affective states have not yet been considered within this research area. Directly linking affective states with cognitive load factors and investigating cognitive load aspects of affective computing could broaden the focus of research in both cognitive load theory and affective computing.

As a step in this direction, this paper reviews basic assumptions of cognitive load theory and major consequences of these assumptions for the design of information presentations, as well as their general implications for affective computing. A recently proposed evolutionary perspective on human cognitive architecture is essential in this analysis, as it allows a broader association of affective computing with an interface between two types of information processing systems - natural (specifically, human) and artificial intelligent information processing systems.

The Architecture of Natural Information Processing Systems

In its basic underpinning assumptions, cognitive load theory relies on the analogy between the information processing aspects of evolution by natural selection and human cognition (Sweller & Sweller, 2006). It considers both biological evolution and human cognition as examples of a broader class of natural information processing systems. It is assumed that operation of such systems is based on the following fundamental principles (Sweller, 2003; for an overview, see Sweller et al., 2011):

1. **The information store principle**: all natural information processing systems include stores of information that govern their activities. In human

cognitive architecture, long-term memory provides this function.

2. **The borrowing and reorganizing principle**: most of information in the store is borrowed and reconstructed from other information stores. In case of human cognition, we build most of long-term memory information structures by imitating people, listening to them and reading materials produced by other people.
3. **The randomness as genesis principle**: all principally novel information is acquired by a random generate-and-test process. In the absence of relevant information from other sources, we acquire it during problem solving by using general methods such as means-ends analysis.
4. **The narrow limits of change principle**: there is a mechanism that prevents large and rapid random changes to the information store that could impair its functionality. Human cognitive architecture includes a working memory as our major information processor that is severely limited in capacity and duration when dealing with novel information, thus reducing the risk of damaging long-term memory. We can consciously process no more than a few items at a time for no longer than a few seconds. If these limits are exceeded, information processing could be inhibited.
5. **The environmental organizing and linking principle**: when information from the store is guiding specific environmentally appropriate activities of the system, the above limits are removed. In human cognition, the severe capacity or duration limits are lifted when organized information from long-term memory is processed in working memory when guiding specific activities (due to the "chunking" or "encapsulation" effect when many connected elements of information are treated as a single unit in working memory)

With the cross-disciplinary approach used in this framework, information should be considered as an attribute of objects of different nature (e.g., genetic codes in biological evolution or written signs in human communication). In general, this attribute of different objects relates to structural patterns in their organization (e.g., Stonier, 1997). Accordingly, information processing is associated with transmitting and adopting such structural patterns. In case of complex natural information processing systems such as evolution of living organisms, these processes are aimed at enhancing their chances of survival. Such complex systems achieve this aim by flexibly adapting to their environments, and all of the above five principles apply to their functioning. The **feedback principle** could be added as a separate one or included in the elaborated last principle in order to explain an essential mechanism of such flexible adaptation.

However, even relatively less complex natural objects (e.g., non-organic matter down to elementary particles) could also be regarded as systems characterized by stable structural patterns. Even subatomic particles maintain their identity, physical integrity and specific behavioral patterns, and engage in information processing. As natural information processing systems, they adhere to, at least, some of the above principles. In accordance with the information store principle, they preserve ("store") specific organizational patterns (e.g., specific patterns of atomic shells). According to the randomness as genesis principle, these patterns may have initially emerged during early stages of cosmic evolution as results of random interactions between particles. Under specific conditions, according to the laws of physics, they formed relatively more stable patterns (achieving a greater time durability provided obvious testing criteria for the random generate-and-test procedure). According to the environmental organizing and linking principle, these information patterns determine interactions of the systems with their environments (e.g., atomic shells determine the nature of atomic interactions in forming more complex molecular structures). However such systems are unable to modify their information structures (learn) and self-reproduce.

As natural information processing systems become more complex, the narrow limits of change principle emerges to assure that this complexity is not destroyed by excessively large random changes. Complex natural systems evolve incrementally with small changes at a time. Also, according to the borrowing and reorganizing principle, for successful survival, complex systems develop the mechanisms of borrowing information from other systems. The complex systems that rely on the whole set of above principles could be treated as *intelligent natural information processing systems* that are flexible and adaptive to unpredictable environmental changes.

Thus, the described general principles of natural information processing systems (Sweller, 2003), including the notion of naturalness, could possibly be traced and extended below the level of biological evolution that provided the essential analogy to the human cognitive architecture (both are intelligent natural information processing systems). This extended framework represents a set of assumptions about the origin of essential features and capabilities of our mind that may possibly be structured in exactly the same way as information processing systems in the rest of the nature.

Human Cognitive Architecture

According to the above framework, human cognitive architecture includes two essential components that define how we process information and learn. One of these components is long-term memory representing our knowledge base (information store) with effectively unlimited capacity and duration. Most of organized knowledge is stored in the form of schemas - generic knowledge structures that are used to mentally categorize and represent concepts and procedures, and govern our behavior. With sufficient practice, schemas can be automated and used without conscious control.

Another essential component of our cognitive system represents a mechanism that limits the scope of immediate simultaneous changes to the knowledge base. This mechanism is associated with the concept of working memory as a conscious processor of information within the focus of attention. Working memory is severely limited in capacity and duration when dealing with novel information (Cowan, 2001). Most of contemporary models of working memory include separate limited processing channels for visual and auditory information (e.g., Baddeley, 1986). Processing limitations of working memory are responsible for learner cognitive overload and profoundly influence the effectiveness of instruction.

When considering human information processing system, cognitive load theory also makes an additional assumption based on evolutionary educational psychology (Geary, 2007). It distinguishes between two major types of information – biologically primary and secondary. It is believed that we are biologically predisposed or "hardwired" to acquire primary information in a rapid, effortless, and mostly automatic way without much of conscious processing and associated cognitive load. For example, information related to speaking and listening native language, engaging in common social interactions, or applying general problem solving strategies belongs to biologically primary information. It has been essential for survival of humans throughout the history of evolution.

On the other hand, biologically secondary information has emerged over a relatively short historical period as a result of cultural and technological developments (e.g., information related to writing and reading skills, science and technology). Acquisition of such information is always effortful and requires conscious controlled processing in working memory. Cognitive load theory applies mostly to the processes of acquisition of biologically secondary information that always generates working memory load.

In most cases, when dealing with novel biologically secondary information, no learning occurs without a cognitive load associated with effortful conscious processing of essential interacting elements of information in working memory. This essential (necessary, productive, and useful) load that is required for achieving specific learning goals is called *intrinsic cognitive load*. Intrinsic cognitive load is caused by specific cognitive activities resulting in new or modified knowledge structures in long-term memory. Such activities involve concurrent processing of interacting elements of information in working memory and integrating them with available knowledge structures in accordance with specific learning goals. Intrinsic load depends on internal complexity of the learning task and is always relative to the level of learner expertise, since what is complex for novices may be simple for experts. In order to achieve meaningful learning outcomes, it is necessary to accommodate this load without exceeding available working memory resources.

In contrast to intrinsic load, *extraneous cognitive load* is a non-essential, unnecessary, and wasteful type of load caused by cognitive activities and processes that are irrelevant to learning goals. This type of load is usually caused by poor instructional design, for example, suboptimal presentation formats or inappropriate selection and sequencing of learning activities with inadequate levels of instructional support. It should be noted that the difference between extraneous and intrinsic cognitive load could be relative to levels of learner expertise: some parts of cognitive load that are essential (intrinsic) for novice learners could become extraneous (irrelevant) for relatively more experienced learners, and vice versa.

Working memory resources that are actually devoted to dealing with intrinsic cognitive load and lead to meaningful learning are defined as *germane resources* (usually, though not quite correctly, called germane cognitive load) in contrast to extraneous working memory resources that are devoted to dealing with extraneous cognitive load (Sweller, 2010). This separate from cognitive load dimension of actually allocated working memory resources stresses the role of germane resources in learner engagement in processing relevant aspects of a task and, therefore, the importance of instructional methods that motivate and engage students in learning-relevant cognitive activities. More engaged and motivated learners invest more of their working memory resources into dealing with intrinsic load thus leading to better learning (Schnotz, 2010). The actual working memory resources invested in learning activities would depend on levels of motivation, attitudes, and affective states of the learner.

Together, the combined intrinsic and extraneous cognitive load determines the total cognitive load imposed on the learner by the learning task. This load determines working memory resources required for processing all the involved elements of information and achieving learning goals by a fully engaged learner. However, it does not necessarily determine actually allocated working memory resources by a specific learner in a specific learning situation. The amount of actually devoted working memory resources depends on how well and fully the learner is engaged in the learning environment.

Sources of Extraneous Cognitive Load

According to cognitive load theory, effective and efficient instruction insures that learner working memory load during learning is kept within its capacity limits. This means that sufficient capacity of working memory is available for processing interacting elements of information that define the information pattern to be learned and thus are essential for achieving specific learning goals (intrinsic cognitive load)

However, most of the instructional techniques developed by cognitive load theory are aimed at eliminating or reducing learner cognitive activities that are not essential for learning and generate extraneous cognitive load. Extraneous cognitive load is typically caused by inappropriate instructional formats or procedures that introduce unnecessary interacting elements of information, for example, performing search-and-match operations that are not relevant to learning, or processing redundant information. Extraneous load is imposed by cognitive activities that a learner is involved in because of the way the learning tasks are organized and presented, rather than because the load is essential for achieving instructional goals. There are following general types of sources of extraneous cognitive load:

1. *Split-attention situations* occur when interacting textual and graphical elements are separated over distance (presented at different locations) or time (presented at different times, non-concurrently). Mental integration of these sources of information might require intense search processes and recall of some elements until other elements are attended and processed. Such processes might significantly increase cognitive load.
2. *Redundancy situations* occur when two or more sources of information can be understood independently without the need for mental integration for example, when text simply re-describes a diagram that can be fully understood without the text. Under these conditions, processing the text and mentally integrating it with the diagram may result in an extraneous cognitive load. The most common form of redundancy is presenting the same information in different modalities, for example, presenting the text in both spoken and written form.
3. *Transiency situations* occur when information disappears before the learner has time to adequately process it. In such situations, learners would need to hold this transient information in their working memory in order to integrate it with the forthcoming pieces. For example, lengthy fragments of continuous spoken text or animations could create such situations.
4. *Advanced learners situations* occur when levels of learner knowledge in a specific area increase to the level when the provision of detailed information may become redundant and distract learners from fluently executing already learned procedures and taking the full advantage of their available knowledge. For instance, providing fully worked out examples of problem solving steps to advanced learners can generate extraneous cognitive load. Since an element or a chunk of information processed in working memory of a particular learner is determined by the schemas the learner holds in long-term memory, with the development of expertise, the size of the learner's chunks increases: many interacting elements for a novice become encapsulated into a single element for an expert. Therefore, the experienced working memory load always depends on levels of learner expertise in a specific task domain.
5. *Inadequate prior knowledge situations* occur when learners do not have adequate knowledge to process new information without cognitive overload. In the absence of relevant knowledge, learners would need to resort to general problem solving strategies such as means-ends analysis to handle novel situations. Such search-based goal-oriented methods always generate excessive levels of extraneous cognitive load that leaves no working memory resources for learning meaningful solution schemas. For example, unguided exploratory learning environments could impose excessive levels of extraneous cognitive load on novice learners.

Managing Cognitive Load to Prevent Negative Affective States

An essential aim of affective computing is creating effective interface between artificial and natural (human) intelligent information processing systems by capturing and interpreting specific affective states of the latter by the former. Negative emotional states could be caused by situations of cognitive overload leading to learner frustration and dissatisfaction with learning activities and own performance. Based on the above potential sources of cognitive overload, it is possible to determine specific factors that may contribute to negative affective states.

Learner working memory could be overloaded if the combined intrinsic (useful) and extraneous (wasteful) cognitive load exceeds its capacity. In this situation, the first step in dealing with cognitive overload is eliminating or reducing the influence of sources of extraneous (wasteful or unnecessary) cognitive load. The following methods are recommended by cognitive load theory to be used in situations of high extraneous cognitive load:

split-attention situations: physically integrating sources of information that are separated in space or time (e.g., embedding verbal information into diagrams or other textual fragments; synchronizing related sources of information in time), using different modalities (auditory and visual) for presenting verbal and pictorial information;

redundancy situations: eliminating redundant sources of information rather than presenting them (e.g., using spoken only text to explain visualizations rather than both spoken and written text simultaneously);

t*ransiency situations*: segmenting lengthy segments of continuous spoken text or animations into smaller portions, or pre-training learners in relevant prior knowledge;

a*dvanced learners situations*: using minimal or reduced levels of non-redundant instructional guidance (e.g., problem solving tasks, exploratory learning environments, completion tasks);

inadequate prior knowledge situations: using well-guided instructions that substitute for missing knowledge structures (e.g., worked examples or explanatory feedback).

If reducing extraneous load still does not eliminate the overload, then additional methods for reducing intrinsic cognitive load should be applied. Among the recommended methods are segmenting the original task into simpler subtasks, pre-training learners in relevant components of knowledge, artificially reducing the number of interacting elements of information during the initial phase of learning followed by the fully interactive element instruction.

Enhancing Positive Affective States

Cognitive means for generating positive affective states are related to optimizing the relation between the learning tasks, learner prior knowledge, and provided instructional guidance. If the provided support is sufficient for learners to make sense of the task on their own, they would more likely experience positive affective states. Also, when motivated by consciously chosen goals, learners experience a sense of control. Learning goals represent an important part of learner knowledge base that performs a guiding role in cognitive processing. Balancing external guidance with learner internal knowledge and goal structures is important for creating positive affective states and higher levels of motivation.

Knowledge base available in learner long-term memory may effectively reduce limitations of working memory by encapsulating many elements of information into higher-level units that are treated as elements in working memory (Ericsson & Kintsch, 1995). Another way of reducing cognitive load is practicing available skills until they can operate under automatic rather than controlled processing (Shiffrin & Schneider, 1977) and allow learners to devote working memory resources to handling more complex situations without cognitive overload.

Thus, more knowledgeable learners can use their available knowledge structures for managing cognitive load. However, if task-relevant knowledge structures are not available in long-term memory, learners have to simultaneously process many new elements of information in working memory resulting in a cognitive overload. Appropriate external guidance may be required to assist these learners in acquiring new knowledge in a cognitively efficient and non-stressful manner. In the absence of a relevant knowledge base or external guidance, the learners may need to resort to weak problem-solving methods based on random search processes that often result in excessive levels of cognitive load with minimal learning.

If, on the other hand, detailed instructional guidance is provided to more experienced learners who have an adequate knowledge base in long-term memory for dealing with the learning task, these learners would need to relate and reconcile the corresponding components of their knowledge and external information. Such co-referencing processes may cause additional cognitive load that would inevitably reduce working memory resources available for learning (e.g., making appropriate generalizations or further strengthening and automating schemas). Thus, as levels of learner expertise increase, relative effectiveness of learning tasks with different levels of instructional guidance may reverse (an expertise reversal effect; Kalyuga, 2007). Presenting more knowledgeable learners with detailed external instructional guidance may inhibit their learning relative to the outcomes that could be achieved with minimal guidance.

Thus, cognitive load consequences of using different learning formats and procedures depend on levels of learner expertise and may result in different affective states. If challenges of the task significantly exceed the available learner knowledge base, the task could cause cognitive overload and emotional unease. On the other hand, when these challenges are too low relative to the available knowledge and skills, the task could be easy and boring, with corresponding affective consequences for the learner. A well fitted learning task that provides challenges just above the level of learner available knowledge base could provide the best motivating power and emotional state. Both unguided effortful search for solutions by novice learners and allocating unnecessary attention to information that could otherwise be processed automatically and effortlessly by more experienced learners would reduce cognitive resources available for meaningful learning. Such unnecessary diversion of attention may emotionally upset and de-motivate learners.

A major instructional implication of the expertise reversal effect and its affective consequences is the need to adapt dynamically instructional formats and levels of instructional guidance to current levels of learner expertise. Learner knowledge needs to be dynamically monitored as it gradually changes during learning, and instructional procedures adjusted accordingly. For example, direct instruction methods could be gradually replaced with less-guided exploratory or problem-based environments as levels of learner expertise increase. Intelligent tutoring sysms (ITS) select learning tasks based on a continuously updating production-rule model of the student (Koedinger & Corbett, 2006) or tracking the cognitive states of the learner by analyzing the content of the dialogue history (Graesser, VanLehn, Rose, Jordan, & Harter, 2001). An important part of such adaptive learning environments is the set of assessment tools that are able to diagnose levels of learner expertise rapidly and in real time.

In addition to assessing acquisition of specific production rules in ITS, a possible method for rapid assessment of expertise could be based on observing how learners approach briefly presented tasks. For example, in the first-step method, learners are asked to rapidly indicate their first step towards the solution of a task presented for a limited time. Experts use their learned solution schemas to rapidly indicate more advanced steps of the solution as their first steps. On the other hand, less experienced learners may only generate a very first immediate move according to the detailed procedure they have learned. Novices could at best be able to indicate only their first attempt in using trial-and-error or means-ends analysis approaches (Kalyuga, 2003). In an alternative rapid verification method, students are briefly presented with potential solution steps at various stages of advancement and asked to rapidly verify their correctness. These diagnostic methods could be effectively used in adaptive computer-based learning environments (Kalyuga, 2006).

Conclusion

There is a close relationship between affective and cognitive states associated with the operation of working memory. Cognitive load caused by suboptimal formatting and sequencing of information presentations or levels of provided instructional guidance that are not tailored to learner prior knowledge could cause negative affective states. Tailoring learning tasks and activities to specific levels of knowledge of individual learners could enhance their positive affective states and increase learning motivation. However, cognitive load theory has not yet considered affective and motivational factors in learning beyond some preliminary ideas (Paas, Tuovinen, van Merrienboer, & Darabi, 2005). The procedures for measuring cognitive load used in this field are mostly based on subjective rating scales with rare exceptions such as the dual-task method (based on monitoring learner performance on a simple secondary task) and some laboratory-based psychophysiological techniques that are not suitable for realistic settings (see Paas, Tuovinen, Tabbers, & van Gerven, 2003; Sweller et al., 2011, for an overview of measurement methods). Some online measures of cognitive load could potentially be used for diagnosing negative affective states.

The inclusion of affective and motivational factors in cognitive load research remains an essential direction for future studies in this area. Establishing connections between affective variables and cognitive load factors, and using methods of affective computing for identifying specific situations of cognitive overload could enhance effectiveness and efficiency of learning environments.

References

Baddeley, A. D. (1986). Working Memory. New York: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences., 24, 87-114.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. Psychological Review, 102, 211-245.

Geary, D. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In J. S. Carlson & J. R. Levin (Eds.), Psychological perspectives on contemporary educational issues (pp. 1-99). Greenwich, CT: Information Age Publishing.

Graesser, A.C., VanLehn, K., Rose, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, 22, 39–51.

Kalyuga, S. (2003). Rapid assessment of learners' knowledge in adaptive learning environments. In U. Hoppe, F. Verdejo, & J. Kay (Eds.). Artificial Intelligence in Education: Shaping the future of learning through intelligent technologies (pp.167-174). Amsterdam: IOS Press.

Kalyuga, S. (2006). Assessment of learners' organized knowledge structures in adaptive learning environments. Applied Cognitive Psychology., 20, 333-342.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. Educational Psychology Review., 19, 509-539.

Kensinger, E. A., & Corkin, S., (2003). Effect of negative emotional content on working memory and long-term memory. Emotion, 4, 378-393.

Koedinger, K. R., & Corbett, A. (2006), "Cognitive Tutors: Technology bringing learning science to the classroom", in Sawyer, K., *The Cambridge Handbook of the Learning Sciences*, Cambridge University Press, pp. 61–78

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist, 38, 63-71.

Paas, F., Tuovinen, J. E., van Merrienboer, J. J. G., & Darabi, A. A. (2005). A motivational perspctive on the relation between mental effort and performance. Educational Technology Research and Development, 53, 25-34.

Schnotz, W. (2010). Reanalyzing the expertise reversal effect. Instructional Science, 38, 315-323.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Psychological Review, 84, 127-190.

Stonier, T. (1997). Information and meaning: An evolutionary perspective. London: Springer.

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), The psychology of learning and motivation (Vol. 43, pp. 215-266). San Diego: Academic 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., & Sweller, S. (2006). Natural information processing systems. Evolutionary Psychology, 4, 434-458.