31

Human Cognitive Architecture

John Sweller
University of New South Wales, Sydney, Australia

CONTENTS

ABSTRACT

Cognitive load theory integrates the origins of human cognition in evolutionary theory with the structures and functions of human cognitive architecture to provide effective instructional design principles. Many of those principles are directly relevant to instructional technology. This chapter outlines the evolutionary bases for human cognitive architecture, examines those aspects of human cognition that are directly relevant to instruction, and discusses the various instructional principles generated by cognitive load theory with specific reference to their applicability to instructional technology.

KEYWORDS

Cognitive load theory: An instructional design theory based on our knowledge of human cognitive architecture.
Human cognitive architecture: The manner in which structures and functions required for human cognitive processes are organized.
Long-term memory: The store holding all knowledge acquired during the processes of learning.
Natural information processing system: The procedures by which natural systems such as human cognition and evolution by natural selection process information.
Working memory: The structure that processes information coming from either the environment or long-term memory and that transfers learned information for storage in long-term memory.

INTRODUCTION

The extent to which any instruction is effective depends heavily on whether it takes the characteristics of human cognition into account. To determine the conditions that maximize learning, we need to closely study human cognition. Once we have established the mechanisms of human cognition, including why those mechanisms have their particular characteristics, we are in a position to design learning environments in accord with human cognitive architecture. Ideal learning environments in accord with human cognitive architecture are not always in accord with realistic learning environments that mimic the real world. Indeed, if human cognition was organized in a manner that always favored learning in realistic over artificial environments, there would never be a need for specialized instructional procedures or even for educational institutions. Observing and interacting with the world would provide the best educational environment. Under such conditions, all instruction would have to be as realistic as possible and would only be required for economic or safety reasons rather than for educational reasons. In contrast, it is suggested in this chapter that, because of the characteristics of human cognition, in many conditions learning is best facilitated by instruction that does not accurately model reality.

Cognitive load theory can be used to determine some of the characteristics of effective instruction. This integrated theory is intended to provide a systematically organized hierarchy leading from evolutionary/biological reasons for the characteristics of human cognitive architecture to the instructional consequences that flow from that architecture. Only some of the instructional procedures generated by cognitive load theory are relevant to educational technology, and those procedures are emphasized in this chapter. We begin by discussing human cognitive architecture within an evolutionary and biological framework.

HUMAN COGNITIVE ARCHITECTURE

Human cognition has evolved to assimilate, process, and use information (or knowledge, used synonymously in this chapter when dealing with human cognition) to direct human action. It constitutes an example of a natural information processing system that is a class of information processing systems that can be found in nature. As a natural information processing system, human cognition is hardly likely to be unique and, indeed, evolution by natural selection can itself be classed as a natural information processing system (Sweller, 2003, 2004; Sweller and Sweller, 2006). The characteristics of such systems will vary depending on their functions, but all natural information processing systems share an identical basic structure or framework. That basic structure in turn can be used to determine how humans deal with information and what types of instructional procedures, including technology-dependent instruction, are likely to be effective.

The essential characteristics of a natural information processing system include: (1) a very large store of information that allows the system to function in the varied environments faced by natural information processing systems; (2) procedures for perpetuating the store of information by transferring information from one entity to another; (3) procedures for changing the store by creating new information to deal with a changing environment; (4) procedures to ensure that changes to the store do not destroy its effectiveness; and (5) procedures to relate information to the external world. These core characteristics as applied to human cognition are discussed here in terms of five principles (Sweller and Sweller, 2006).

The Information Store Principle and Human Long-Term Memory

Human long-term memory provides the human cognitive system with the large store of information indispensable to a natural information system. A species' genome has a similar function in biological evolution. Our knowledge of the function of long-term memory has altered over time, and, arguably, specifying the role of long-term memory constitutes the primary finding of the cognitive science revolution. We no longer see long-term memory as a repository of isolated, unrelated facts that are occasionally stored and retrieved; instead, it is the central structure of human cognitive architecture.

Origins, Evidence, and Implications

De Groot's (1946/1965) work on expertise in the game of chess can be seen as the initial work transforming our perception of the role of long-term memory. De Groot found that, if chess masters are shown a board configuration taken from a real game for about 5 seconds, they can reproduce it much more accurately than weekend players. Chase and Simon (1973) reproduced this finding but also found that masters and weekend players did not differ in their ability to reproduce random board configurations. This difference between experts and novices in memory for real configurations and situations has been replicated in a variety of fields (Egan and Schwartz, 1979; Jeffries et al., 1981; Sweller and Cooper, 1985). Furthermore, it is the only reliable difference that has been obtained differentiating novices and experts in problem-solving skill and is the only difference required to fully explain why an individual is an expert in solving particular classes of problems. A chess master has learned to recognize many thousands of board configurations. When faced with a configuration, he or she recognizes it and knows the best move to make given that configuration. We all have that skill in our own areas of expertise and, indeed, the ability to recognize situations and the appropriate actions that they require constitutes our skill base in a given area.

Several implications flow from these findings. First, problem-solving skill is domain specific. A brilliant mathematician has acquired mathematical problem-solving skills that are unlikely to transfer to, for example, financial or personal relationship skills. The skills allow a mathematician to recognize mathematical problem states and the best moves associated with them. Such skills are only useful in a mathematical context. Second, the work on novice–expert differences explains why it takes so long to become an expert in a substantial field. It takes about 10 years of concentrated work for a person to become a chess grand master (Simon and Gilmartin, 1973). During that time, the person is not learning complicated, general problem-solving strategies. There is no evidence that learnable or teachable general problem-solving strategies exist; rather, the chess player is learning to recognize the huge number of board configurations required for chess expertise, just as all of us are required to learn the huge number of problem situations required to acquire problem-solving skill in a particular area. Third, data on novice–expert differences demonstrate the central importance of knowledge or information held in long-term memory to skill in any area (Chi et al., 1982). To be skillful, we must have a huge stock of knowledge available. Based on this conceptualization, long-term memory is critical to skilled performance. Knowledge held in long-term memory allows us to function in the variety of contexts in which we find ourselves in much the same way as a genome allows a species to function biologically. Accordingly, a major function of instruction, including technology-based instruction, is to ensure that appropriate knowledge is held in long-term memory.

The Borrowing Principle and Transferring Knowledge

Once discovered, information held by natural information processing systems must be perpetuated, and a primary function of education is to ensure that knowledge is not lost because, as will be discussed below, there are structural reasons why the act of discovery is inordinately difficult. In genetics, the store of DNA-based information is perpetuated over long periods of time by sexual and asexual reproduction. Asexual reproduction includes precise copying from one generation to the next and appears to have no cognitive equivalent. In contrast, sexual reproduction is a constructive procedure designed to ensure that offspring differ from their parents. This process has an inevitable random component, with the precise combination of genetic material from male and female ancestors being intrinsically unpredictable.

Psychological mechanisms are required equally to preserve information held in long-term memory by its transmission from individual to individual. Accordingly, we have evolved to both efficiently transmit and receive information from other humans in either auditory or visual form. We imitate what others do, listen to what they say, and read what they write. That skill permits knowledge held in an individual's long-term memory to be perpetuated indefinitely. Nevertheless, as is the case with sexual reproduction, the process is constructive in nature. We combine new information with information previously stored in long-term memory (Bartlett, 1932). As is also the case with sexual reproduction, the process has a random component in that we cannot predict precisely how the two sources of information will be combined. This randomness has implications for the narrow limits of change principle discussed below.

Origins, Evidence, and Implications

A major function of instruction is to organize efficient procedures that will permit knowledge to be transferred to the long-term memories of learners. Cognitive load theory has generated a variety of instructional procedures relevant to technology-based instruction (see below). All are dependent on the borrowing principle. The success of cognitive load theory in generating instructional procedures is, at least in part, due to its emphasis on learning from presented information. That success provides some of the evidence for the borrowing principle.

In earlier eras, the suggestion that a function of education is to transfer knowledge would have been considered self-evident. More recently, an emphasis on discovery and constructivist procedures encouraged learners to *discover* knowledge rather than to have instructors *transmit* knowledge. The impossibility of discovering even a tiny fraction of the huge amount of information required in the modern world was partially obscured by the field's ignorance of the massive size of long-term memory. That ignorance is being rectified and increasing numbers of investigators are reacting against the previous orthodoxy (Kirschner et al., 2006; Klahr and Nigam, 2004; Mayer, 2004).

Our increasing knowledge of the importance of imitation in human learning provides additional evidence for the borrowing principle. The discovery of mirror neurons that fire in the same manner when we take an action, observe someone else make the same action, or even listen to a sentence describing the action provides neuropsychological evidence for the importance of imitation as a learning mechanism (Tettamanti et al., 2005).

Randomness as Genesis Principle and Creating Knowledge

The need to transfer knowledge should not obscure the fact that knowledge must first be created to have something to transfer. In addition, circumstances change and current knowledge may no longer be adequate. The procedure for changing the store of information held by natural information processing systems is standard. All creative changes are random, but only effective changes are retained, and ineffective changes are jettisoned. In evolution by natural selection, this mechanism is well accepted. Random mutation is the initial source of all genetic variation with only successful mutations resulting in reproduction and continuation. In psychology, and especially instructional psychology, the suggestion requires further explanation.

Random generation followed by effectiveness testing is unavoidable in functioning, natural information-processing systems. Consider a person solving a novel problem, an activity intended to create a solution that is new for that person. When deciding on a problem-solving move or a series of moves, the two basic categories of move generators are random generation and generation based on previous knowledge. These two categories or a combination of these two categories are the only source of move generation. Previously acquired knowledge held in long-term memory can act as a central executive determining moves, but, of course, that knowledge is not always available. There is no general central executive that can determine moves in the absence of relevant knowledge. Random generation, either mentally or in the real world, is left as the only other alternative.

Origins, Evidence, and Implications

The origin of this principle is largely logical rather than empirical. If knowledge to determine a problem-solving move is unavailable, random generation followed by tests for effectiveness is the only other option. We can see the impact of this logic in computer models of problem solving. Such models (Sweller, 1988) randomly generate moves if information is lacking.

Does random generation mean that whether or not a novel problem is solved is entirely due to chance? Random generation can only function properly when coupled with effectiveness testing, and the combination of random generation and effectiveness testing provides the knowledge- or information-creating process of natural information-processing systems. By only accepting effective alterations to an information store and rejecting ineffective alterations, the effectiveness of the store can increase incrementally over time. By transferring the information in a store to other entities that can continue the process of random generation followed by effectiveness testing indefinitely, very complex, sophisticated stores can by built by natural information-processing systems. Both evolution by natural selection and human cognition provide examples. Nevertheless, it must be recognized that random generation followed by effectiveness testing is a very slow way of generating knowledge (Cooper and Sweller, 1987; Sweller and Cooper, 1985) and should only be used when the alternative procedure of knowledge transfer is unavailable. Knowledge transfer is vastly more effective. Accordingly, technology-based instruction requiring problem solving (random generation followed by effectiveness testing) can be expected to be less effective than instruction providing demonstrations (knowledge transfer via the borrowing principle).

The Narrow Limits of Change and Human Working Memory

There are limits to knowledge creation. In the initial instance, knowledge creation occurs using the randomness as genesis principle, which requires random generation followed by effectiveness testing. Knowledge may be altered using the borrowing principle but that also involves random generation when new information is combined with old information. Mechanisms are required to ensure that any alterations to the store of information in long-term memory maximize the probability that a particular change will be effective and minimize the probability that a change will destroy the functionality of the store. These conditions are met by the use of a series of small, incremental changes, each tested for effectiveness, rather than a single, very large change.

Assume that the effectiveness is being tested of adding a permutation of 4 elements to the store. Assume further that no knowledge is available to eliminate some of the possible permutations or select others for priority testing. Under these circumstances, there are $4! = 24$ possible permutations. In contrast, assume that rather than 4 elements being under consideration there are 10. The number of possible permutations is $10! = 3{,}628{,}800$. If only a limited number or perhaps only one permutation is actually effective, dealing with more than a very small number of elements that must be combined randomly and tested for effectiveness is futile because the presence of any more elements rapidly leads to a combinatorial explosion that no natural information processing system can possibly handle unless it has prior knowledge.

Origins, Evidence, and Implications

This factor alone explains some of the major characteristics of human cognitive architecture. When dealing with novel information for which there is no or limited prior knowledge, human working memory is extremely limited in capacity (Miller, 1956). Although the precise nature of this limitation has provided a source of research and discussion (Cowan, 2005), few would dispute the existence of the limitation. The consequences of this limitation have profound consequences for instructional design. Instructional designs that ignore this limitation are likely to be ineffective.

The Environment Organizing and Linking Principle

This principle pertains to how we use information to function in our environment. In contrast to the limitations of working memory, when dealing with familiar information that is already organized in long-term memory, there is no functional reason for working memory to be limited. Accordingly, huge amounts of organized information can be transferred from long-term memory to working memory without overloading working memory (Ericsson and Kintsch, 1995). That information can then be used to allow us to function in our complex environment; for example, if readers attempt to reproduce from memory the immensely complex set of squiggles that constitute the last sentence, most will be able to do so effortlessly. We are able to do so because of the organized, schematic information held in long-term memory that can be transferred and used in working memory. That information allows us to read, process, organize, and relate text to the external world. Similarly, whereas changes to a genome must be small and incremental, huge amounts of previously organized genetic information can be used simultaneously to produce the complex proteins required for biological survival in a complex environment.

Origins, Evidence, and Implications

This principle has multiple origins. Miller's (1956) evidence that multiple elements could be "chunked" together to act as a single element in working memory, schema theorists demonstrating that the manner in which we process information depends on previous knowledge (Bartlett, 1932), and, more recently, Ericsson and Kintsch's (1995) concept of long-term working memory all contribute to the environment organizing and linking principle. Such lines of research demonstrate how we use our large store of information to impose order and meaning on our environment. The environment organizing and linking principle provides the ultimate justification for human cognition. The purpose of the previous four principles is to permit cognition to occur so we can function mentally in our environment.

COGNITIVE LOAD THEORY

These characteristics of human cognitive architecture have direct implications for instructional design and hence the design and purpose of technology-based instruction. The preferred characteristics of technology-based instruction can be determined by using cognitive load theory (Clark et al., 2006; Paas et al., 2003, 2004; Sweller, 2005a; van Merriënboer and Sweller, 2005), as well as closely related theories such as Mayer's theory of multimedia learning (Mayer, 2005) or van Merriënboer's (1997) 4C/ID model, which use this architecture to determine general instructional design principles. As might be expected, all theories that use this architecture are compatible and make similar or identical predications.

Cognitive load theory specifies two sources of cognitive (or working memory) load that determine the effectiveness of instruction. Extraneous cognitive load is due to inappropriate instructional designs and so must be reduced. If working memory resources are being fully expended, a reduction in extraneous cognitive load is required to permit an increase in germane cognitive load, a form of cognitive load that can result in useful alterations to long-term memory. These two sources of cognitive load determine the effectiveness of instruction, but a third source of cognitive load, intrinsic cognitive load, cannot be manipulated without compromising understanding. Intrinsic cognitive load (Sweller, 1994) can be thought of as the intrinsic complexity of the material being studied. For learners at a given level of expertise, that complexity can be reduced (Pollock et al., 2002) but only by reducing learners' understanding of the subject matter. Intrinsic cognitive load can, of course, also be reduced by learning (i.e., by testing learners with more expertise).

These sources of cognitive load are additive and cannot exceed the available capacity of working memory. If intrinsic cognitive load is low, it may be possible for germane cognitive load to be high even with inappropriate instructional techniques causing a high extraneous cognitive load. The low intrinsic cognitive load is likely to leave sufficient working memory resources for students to learn even with a poor instructional design. In contrast, if intrinsic cognitive load is high due to high complexity material, unless the extraneous cognitive load is low there may be insufficient working memory capacity to permit a level of germane cognitive load that can result in learning. With a high intrinsic cognitive load, it is essential to keep the extraneous cognitive load low to permit a sufficient level of germane cognitive load. In other words, instructional design becomes critical with complex material.

Based on the cognitive architecture outlined above, extraneous cognitive load can be reduced and germane cognitive load increased by assisting learners to transfer information to long-term memory. Novel information coming from the senses is not organized and imposes a heavy load on working memory (the narrow limits of change principle). Information coming from long-term memory is organized and imposes a minimal load on working memory (the environment organising and linking principle). In other words, once learned, information no longer imposes a working memory load. How should we assist learners to transfer novel information to long-term memory? Wherever possible, that information should be in an organized form so learners do not have to expend working memory resources in imposing an organizational structure. Almost all information that learners must acquire has previously been laboriously organized over many generations using the randomness as genesis principle outlined above. Nothing is gained by requiring learners in an instructional setting to attempt to use random generation followed by tests of effectiveness rather than the borrowing principle.

Contrary to much current educational dogma (Kirschner et al., 2006; Mayer, 2004) and as indicated above, our cognitive architecture strongly facilitates learning through knowledge transfer via the borrowing principle rather than knowledge generation via the randomness as genesis principle. Not only should knowledge be presented to learners rather than have them engage in the impossible task of attempting to generate it themselves, but it should also be presented in a manner that reduces extraneous cognitive load and maximizes germane cognitive load. Cognitive load theory, making use of the borrowing principle, has generated a range of techniques intended to achieve this purpose. Most of those techniques are relevant to technology-based instruction and will be discussed next. Each technique has been studied as an experimental effect using randomized, controlled experiments in which an instructional technique generated by cognitive load theory is compared to an alternative, usually more traditional technique. These effects provide the instructional recommendations generated by cognitive load theory.

Cognitive Load Theory Effects and Technology-Based Instruction

Worked Example Effect

The worked example effect (Sweller and Cooper, 1985) occurs when novice learners studying worked solutions to problems perform better on a problem-solving test than learners who have been given the equivalent problems to solve during training. This effect flows directly from the cognitive architecture described above. When solving an unfamiliar problem, problem solvers have no choice but to use a random generation followed by effectiveness testing procedure at those choice points where they have insufficient knowledge to direct choice. The working memory load associated with using an effective problem-solving procedure will interfere with learning (Cooper and Sweller, 1987; Paas and van Merriënboer, 1994) and so constitutes an extraneous cognitive load. In contrast, using the borrowing rather than the randomness as genesis principle should enhance learning. Showing learners how to solve problems via worked examples by providing an organized structure rather than leaving learners to devise their own organization using their limited capacity working memory should minimize extraneous cognitive load. Findings from a large number of studies carried out by many researchers in the 1980s and 1990s provided strong support for these hypotheses.

There is every reason to suppose that the worked example effect should apply directly to technology-based instruction. Consider a computer-based simulation. A simulation should demonstrate a process or procedure. It should not require learners to solve novel problems even if problem solving is part of the process being simulated. The aim of a simulation demonstrating problem solving should be identical to the aim of any other instruction in problem solving, and that aim is to assist learners in acquiring knowledge in long-term memory concerning the problem-solving procedures relevant to that particular problem. Once that knowledge is acquired, learners will recognize the problem as belonging to a category requiring particular moves for solution (Chi et al., 1982). Searching for novel solutions using random generation followed by effectiveness testing as part of a simulation is no more likely to achieve this aim than searching for novel solutions in any other instructional context. Using the borrowing principle, a good simulation can eliminate randomness by precisely indicating the structure of what needs to be learned. (Practicing familiar problem solutions has a different function and is discussed below in the section on the expertise reversal effect.)

Split-Attention Effect

Assume that the material being presented to learners consists of two or more sources of information. Assume further, that the sources of information are unintelligible in isolation and can only be understood in conjunction with each other. A geometry worked example provides an instance. The diagram tells novice learners little if anything of the problem solution, and the associated statements are likely to be unintelligible without reference to the diagram. To understand the material, learners must mentally integrate the two sources of information. Mental integration requires working memory resources to be used to search for appropriate references between the multiple sources of information. That search process is indistinguishable from the random generation followed by effectiveness testing of problem solving. All searches not based on knowledge require the randomness as genesis principle. If a geometry solution includes the statement "angle ABC," then learners must randomly choose an angle and test whether it is angle ABC, with the process continuing until the angle is found. Little can be expected to be learned by this use of working memory resources. In contrast, borrowing other people's knowledge can reduce extraneous cognitive load. If the instructions are physically integrated so "angle ABC" is clearly associated with the appropriate angle, obviating the need to search for the angle, then extraneous cognitive load should be reduced and learning enhanced. That result also has been obtained on numerous occasions (Ayres and Sweller, 2005; Sweller et al., 1990). It must be emphasized that the split-attention effect *only* occurs when multiple sources of information must be integrated before they can be understood. Multiple sources of information that can be understood in isolation should not be physically integrated. Different instructional procedures are required when multiple sources of information can be understood in isolation and are discussed in the Redundancy Effect section.

Technology-based instruction that ignores the split-attention effect is likely to be less than effective; for example, a simulation demonstrating the functions of a mechanical device in which the function of a part of the device can only be understood in relation to the function of another part of the device runs the risk of split attention. Whenever possible, the simulation should be structured to clearly indicate the relation between the two parts. In addition, any written text should be formatted in a manner that reduces or eliminates a search for referents. Learners should not be left to work out relations between aspects of a simulation. Working memory resources can be better employed.

Modality Effect

This effect occurs under the same conditions as the split-attention effect in that both effects occur under conditions with multiple sources of information that cannot be understood in isolation and so must be integrated (Low and Sweller, 2005). The effect relies on a particular characteristic of working memory. In the previous discussion of human cognition, working memory was treated as a unitary concept; in fact, it is best thought of as consisting of multiple channels or processors (Baddeley, 1992). A visual processor for dealing with two- or three-dimensional objects and an auditory processor for dealing with language are partially independent. As a consequence, the simultaneous use of both processors can expand the effective size of working memory, under some circumstances. One of those circumstances consists of the conditions leading to the split-attention effect—namely, multiple sources of information that must be integrated before they can be understood. Under those conditions, the use of both the auditory and visual processor can expand the effective size of working memory in an instructionally favorable manner.

Consider a geometry example again as discussed above. To understand a geometry worked example, the learner must simultaneously consider both the diagram and the text because neither conveys the required meaning in isolation. By presenting the verbal information in spoken rather than written form, working memory can be effectively expanded because information is shifted from the overloaded visual processor to be shared by both the visual (for the diagram) and auditory (for the text) processors. According to cognitive load theory, such an expansion of effective working memory should facilitate learning. Tindall-Ford et al. (1997) demonstrated the modality effect when learners presented with instructions in a visual, split-attention format learned less than learners presented with the same material but with all of the verbal information presented in spoken rather than written form.

The modality effect is directly applicable to technology-based instruction. Whereas an aspect of technology-based instruction requires verbal input to be intelligible, extraneous cognitive load can be reduced by using spoken rather than written text. The use of spoken text is particularly important during an animation. The expansion of effective working memory due to the use of both the auditory and visual processors can permit an animation to be viewed while simultaneously attending to speech that explains otherwise unintelligible aspects of the animation. The use of written rather than spoken text runs the risk of overloading the visual processor. In contrast, appropriate use of both visual and auditory information can maximize the potentially powerful effects of the borrowing principle.

Redundancy Effect

The previous two effects discussed, the split-attention and modality effects, apply to multiple sources of information, each of which is unintelligible in isolation. In contrast, the redundancy effect applies to multiple sources of information that are intelligible in isolation. That difference in the logical relation between sources of information results in quite different instructional consequences. The difference is important because frequently, on the surface, conditions that lead to the split-attention or redundancy effects can look identical. It is only by considering the relation between the multiple sources of information that the appropriate instructional recommendations can be provided.

In the current context, redundant information is defined as any information that is not relevant to learning. Most commonly, redundant information consists of the same information presented in different forms or media such as presenting the same verbal information in spoken and written form, but it can also consist of any unnecessary, additional information such as decorative pictures, background sound, or cartoons.

In instructional contexts, redundancy can frequently be found when textual information repeats information found in a diagram. A diagram indicating the flow of blood in the heart, lungs, and body along with a statement that "blood flows from the left ventricle to the aorta" provides an example of redundancy (Chandler and Sweller, 1991). Although a geometry diagram and its associated statements would appear, on the surface, to have the same properties and hence are governed by the same instructional principles as a blood flow diagram and its associated statements, they are structurally very different and require very different formatting. A geometry diagram tells us little of a problem solution and requires the statements for an intelligible solution to be communicated to learners. Those statements must be integrated with the diagram so a search for referents is reduced. In contrast, the blood flow diagram can be fully intelligible in its own right and provide a full explanation. The appropriate instructional procedure is not to integrate the statements with the diagram but rather to eliminate the least effective source of information, which, in this case, is the set of statements. Many experimental examples demonstrate that the elimination of redundancy facilitates learning (Chandler and Sweller, 1991).

Although the redundancy effect can be considered counter-intuitive, from a cognitive load theory perspective the reason why redundant information has negative consequences is straightforward. Attending to unnecessary information and attempting to integrate it with essential information requires working memory resources that consequently are unavailable for learning. Redundant information imposes an extraneous cognitive load. It is an ineffective use of the borrowing principle.

There are many forms of redundancy other than the diagram and text redundancy described above (for a detailed summary, see Sweller, 2005b). All forms of redundancy potentially apply to technology-based instruction; for example, a spoken commentary associated with a simulation should only be included if the visual material is unintelligible without the commentary. If visual material is intelligible in isolation, a spoken commentary will increase cognitive load as learners attempt to integrate the auditory and visual information. When devising a simulation, care must be taken to ensure that all information presented is essential and not simply an alternative way of presenting the same information. Equally damaging is additional information that bears little or no relation to the required information. It is a trap to assume that additional information will, at worst, be neutral in its effect and could be beneficial.

Increasing technical sophistication is permitting increasingly realistic simulations. By definition, the real world is realistic, but we developed instructional systems precisely because many aspects of the real world provide inadequate instruction. That inadequacy frequently is caused by redundancy. A realistic simulation that provides a depiction of a mechanical system may be indistinguishable from the real mechanical system, but that system may be almost useless as an instructional tool. When learning how the blood flows in the heart, lungs, and body, most of the structures, functions, and characteristics of the body are irrelevant, which is why it took so long to discover the processes of the circulatory system using the randomness as genesis principle. The "realistic" but irrelevant features served to conceal the critical features. Realistic features and processes should not be included in a simulation if the sole reason for inclusion is realism. To avoid redundancy and maximize the effect of the borrowing principle, there should be a clear instructional reason for including any information.

Expertise Reversal Effect

Two related points must be made concerning the discussion of the previous effects. First, all of the above effects assume that learners are novices. It is novices who most frequently require instruction via the borrowing principle. Second, the previous explanation of the redundancy effect implicitly assumed that redundancy is purely a function of the materials being used; in fact, it is equally a function of levels of expertise. Information that is redundant for a more expert learner may be critically necessary for a less expert learner. A novice may need to borrow information from someone else, an expert may not. The expertise reversal effect was born from these considerations.

The effect occurs when an instructional procedure that is relatively effective for novices compared to a control procedure first loses its advantage as levels of expertise increase and then begins to be worse than the control procedure with further increases in expertise (Kalyuga et al., 2003). As an instance, studying worked examples is better than solving the equivalent problems for novices, but with increased expertise solving problems becomes better than studying examples (Kalyuga et al., 2001). As other instances, for novices, integrated format or dual-modality instruction is better than split-attention format instruction. With increasing expertise, rather than integrating, for example, diagrams and text or using dual-modality instruction, it is better to eliminate the text entirely (Kalyuga et al., 1998, 2000).

The expertise reversal effect is a complex effect that relies on redundancy. As expertise increases, previously essential information becomes redundant and so imposes an extraneous cognitive load. Studying worked examples may be essential for novices to reduce cognitive load, but such activity becomes redundant with increasing expertise and is better replaced by practice at solving problems which, at higher levels of expertise, no longer imposes a cognitive load. Similarly, explanatory text may be essential for novices and so should be physically integrated with diagrams or presented in spoken form to reduce extraneous cognitive load. With increasing expertise, the text becomes redundant and should be eliminated.

The expertise reversal effect suggests that the detail provided in technology-based instruction should be determined by the knowledge base of the learners. Details that are essential for novices may be redundant for more expert learners. Thus, technology-based instruction must be constructed so its specifications change with changes in expertise. Furthermore, if instructional materials are to change with changes in expertise, a method is required to rapidly determine levels of expertise. Kalyuga and Sweller (2004) provided a rapid test of knowledge based on the cognitive architecture described above. During instruction, learners were presented a partially completed problem and asked to indicate the next step required for solution. The extent to which a learner knows the next step to solution depends on the knowledge base held in long-term memory. That information can be used to determine subsequent instruction. It should similarly be possible to use this rapid assessment technique to determine the nature of any subsequent instruction.

Guidance Fading Effect

The guidance fading effect (Renkl and Atkinson, 2003; Renkl et al., 2004) is closely related to the worked example and expertise reversal effects and is also a compound effect. It occurs when novices are initially presented with worked examples, but with increasing expertise those worked examples are replaced by completion problems (van Merriënboer et al., 2002) in which a partial solution is provided and the learner is required to complete the problem. With further increases in expertise, the completion problems should be replaced by full problems. Again, this sequence is predicated on the assumption that what constitutes an extraneous cognitive load depends not just on the nature of the instruction but on an interaction between the instructional procedures and learner characteristics in the form of levels of expertise. At low levels of expertise, the learner must make heavy use of other people's knowledge via the borrowing principle. With increasing expertise, the same information can be borrowed from the learner's own long-term memory and used for practice purposes.

Technology-based instruction should take account of the guidance fading effect by initially providing substantial guidance which should be gradually faded as expertise increases. As an example, initially, learners should be shown exactly what they need to do with minimal action required on their part. With increases in expertise, determined by rapid assessment techniques, learner activity should be increased and guidance decreased. Ultimately, it should be possible to remove all guidance with the learner simply practicing the skill.

Imagination Effect

The imagination effect occurs when learners who are asked to imagine a procedure or concept learn more than learners who are asked to study the same procedure or concept (Cooper et al., 2001; Leahy and Sweller, 2005). Imagination instructions ask learners to turn away from the material and attempt to imagine the relevant procedures or concepts. Imagining involves running material through working memory which should assist in the transfer of the information to long-term memory. The technique is highly effective but only when used by learners with sufficient experience in the domain to be able to process all of the necessary information in working memory without assistance from the instructional material. For novices, attempts to run a procedure through working memory are likely to fail, so instructions to study material, which involves considering it while looking at it, are superior to imagination instructions. The switch from studying being superior to imagination being superior provides another example of the expertise reversal effect. Again, as was the case for the expertise reversal and guidance fading effects, for novices information is best borrowed from another person's long-term memory, but as levels of expertise increase that information can be borrowed from one's own long-term memory for purposes of practice, in this case mental practice.

The imagination effect provides information on what mental activity learners should be engaged in when dealing with a simulation. Initially, they should simply study or interact with the simulation materials because due to working memory limitations they are unlikely to have sufficient knowledge to be able to effectively imagine the procedures and concepts. With increasing expertise, they should attempt to imagine the information covered by the instruction because that procedure seems to be the most rapid technique for transferring information to long-term memory and so increasing levels of expertise.

Element Interactivity Effect

The expertise reversal effect discussed above places an emphasis on the cognitive load implications of individual differences in expertise. Differences in the structure of the information being considered are equally important. None of the above effects is obtainable using low-complexity material (Sweller, 1994) that has a low intrinsic cognitive load. Recall that total cognitive load is an addition of extraneous, intrinsic, and germane cognitive load. The above effects are primarily determined by an excessive extraneous cognitive load that reduces germane cognitive load because working memory capacity is exceeded. If intrinsic cognitive load is low, a high extraneous cognitive load may not matter a great deal. There may be sufficient working memory capacity available to enable germane cognitive load and its attendant rapid learning to occur.

What determines levels of intrinsic cognitive load? The only relevant factor within a cognitive load theory framework is element interactivity, which is determined by the number of interacting elements that must be considered simultaneously to understand the material. Some information is low in element interactivity in that the elements can be learned one element at a time without considering any other elements. Learning technical terminology provides an example. One can learn the name of a component without learning the names of any other components, so working memory load may be very low. Cognitive load effects are not likely to be relevant when intrinsic cognitive load is low. In contrast, learning how components interact in a machine has high element interactivity because it may be impossible to understand the function of one component without simultaneously considering the function of all of the components. High element interactivity results in a high intrinsic cognitive load, leaving little working memory capacity available for learning. Under these circumstances, levels of extraneous cognitive load become critical, and the cognitive load effects discussed above become relevant; thus, cognitive load effects become critical when technology-based instruction deals with complex, high element interactivity material.

Isolated Interacting Elements Effect

When element interactivity is very high, it may be impossible for learners to understand the material because it may be impossible for them to simultaneously process all of the interacting elements in working memory. How should such material be presented? From a cognitive load theory perspective, the only way seems to be to initially present the material as individual elements ignoring their interactions. This procedure will permit the elements to be learned but without understanding. Once the individual elements have been learned, their interactions can be emphasized. It is only at that point that the material will be understood because it cannot be understood by simply considering individual elements. Empirical work has demonstrated that teaching individual elements first, at the expense of understanding, followed by teaching the interactions between elements results in more effective learning than attempting to have learners understand very high element interactivity material right from the beginning of instruction (Pollock et al., 2002).

The isolated, interacting elements effect may have considerable relevance to technology-based instruction such as instructional simulations. Material is difficult to understand because it is high in element interactivity. Presenting that material in a realistic fashion during a simulation may be condemning learners to attempting to understand information that vastly exceeds their working memory capacity because a realistic simulation may involve a huge number of interacting elements. A less realistic simulation with fewer interacting elements may be more readily understood and learned. Although a full understanding of the material is impossible by this technique, it is equally impossible to process a large number of interacting elements simultaneously. It may be better to provide simulations that result in limited understanding initially, followed by more complete versions that permit full understanding. In this manner, learners may be spared the need to attempt to understand material that is quite impossible for them to deal with.

DISCUSSION

Cognitive load theory provides an integrated system dealing with the evolutionary origins of human cognition leading to an explanation of cognitive structures and processes. In turn, those structures and processes can be used to generate instructional principles. The generation of applications in the form of instructional principles provides some warrant for the validity of the original cognitive architecture. This unified system can provide a base for instructional design including the structure and function of technology-based instruction.

There are two aspects of technology-based education in general and of the work reported in this chapter that require emphasis. First, the use of technology in education should not be based merely on the availability of technology. There is a long history of new technological applications such as radio, films, television, mainframe computers attached to terminals, standalone microcomputers and now the Web being hailed as potentially revolutionizing education. Frequently, these technological advances have had minimal, long-term educational impact. While technology changes, human cognitive architecture does not. The introduction of educational technology without reference to its cognitive consequences is unlikely to be effective. The five natural information processing principles outlined above provide an initial guide. According to those principles, learning consists of changes to the long-term store, the most efficient method of bringing about those changes is by borrowing knowledge from knowledgeable educators, and that knowledge must be structured in a manner that reduces working memory load. I am not aware of any evidence that simply using technology will necessarily conform to any of the five principles. Instruction that ignores human cognitive architecture is likely to be ineffective whether or not it uses technology. In contrast, technology-based instruction that is explicitly structured to conform to what we know of human cognition has a much better chance of being effective.

All of the cognitive load effects discussed above were based on the assumption that the aim of instruction is the acquisition of knowledge in long-term memory and that the best way of achieving this aim is to make use of the borrowing principle by providing direct instructional guidance organized to reduce extraneous working memory load.

The second point that requires emphasis concerns how we determine whether technology-based education is effective. The instructional effects described above were not only based on our knowledge of human cognitive architecture but also were tested using controlled, randomized experimental designs. Experiments in which participants are randomly allocated to two or more experimental groups with different instructional procedures but identical test procedures are essential when determining effective instructional procedures in this field. Verification procedures in which instructors and learners are presented instruction using a new technology and simply asked if they think it is effective or if they enjoyed it have very limited utility.

While cognitive load theory indicates many conditions under which instruction may or may not be effective, those conditions are closely dependent on the cognitive architecture discussed and, specifically, on the characteristics of working and long-term memory. The theory is not intended as a general theory of cognition or of instruction, and so many important variables are not considered by the theory; for example, despite some discussion about including motivational variables in cognitive load theory and considering this issue in future research, at present it has not been incorporated into the theory. That incorporation will, in all probability, only occur if relations between cognitive architecture and motivation can be established.

In conclusion, advances in instructional technology have provided us with the ability to use instructional procedures that until recently were difficult or impossible to implement. Frequently, those procedures have been recommended purely because they are now possible rather than because there was evidence for their cognitive effectiveness or even desirability. Cognitive load theory is intended to provide a guide to the types of techniques that are likely to work. It has been tested using controlled, randomized experiments over many years by many investigators in a wide variety of technology-based environments. There is every reason to suppose that the use of our knowledge of human cognition will continue to generate technology-based instructional procedures.

ACKNOWLEDGMENTS

The work reported here was supported under Office of Naval Research Award Number #N00014-04-1-0209, as administered by the Office of Naval Research. The findings and opinions expressed in this report do not necessarily reflect the positions or policies of the Office of Naval Research.

REFERENCES

Ayres, P. and Sweller, J. (2005). The split-attention principle. In *Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer, pp. 135–136. Cambridge, U.K.: Cambridge University Press.*

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

Bartlett, F. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge, U.K.: Cambridge University Press.

Chandler, P. and Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cogn. Instruct.*, 8, 293–332.

Chase, W. G. and Simon, H. A. (1973). Perception in chess. *Cogn. Psychol.*, 4, 55–81.

Chi, M., Glaser, R., and Rees, E. (1982). Expertise in problem solving. In *Advances in the Psychology of Human Intelligence*, edited by R. Sternberg, pp. 7–75. Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, R., Nguyen, F., and Sweller, J. (2006). *Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load*. San Francisco, CA: Pfeiffer.*

Cooper, G. and Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *J. Educ. Psychol.*, 79, 347–362.

Cooper, G., Tindall-Ford, S., Chandler, P., and Sweller, J. (2001). Learning by imagining. *J. Exp. Psychol. Appl.*, 7, 68–82.

Cowan, N. (2005). *Working Memory Capacity*. Oxford, U.K.: Psychology Press.*

De Groot, A. (1946/1965). *Thought and Choice in Chess*. The Hague: Mouton.

Egan, D. E. and Schwartz B. J. (1979). Chunking in recall of symbolic drawings. *Memory Cogn.*, 7, 149–158.

Ericsson, K. A. and Kintsch, W. (1995). Long-term working memory. *Psychol. Rev.*, 102, 211–245.*

Jeffries, R., Turner, A., Polson, P., and Atwood, M. (1981). Processes involved in designing software. In *Cognitive Skills and Their Acquisition*, edited by J. R. Anderson, pp. 255–283. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kalyuga, S. and Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *J. Educ. Psychol.*, 96, 558–568.

Kalyuga, S., Chandler, P., and Sweller, J. (1998). Levels of expertise and instructional design. *Hum. Factors*, 40, 1–17.

Kalyuga, S., Chandler, P., and Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *J. Educ. Psychol.*, 92, 126–136.

Kalyuga, S., Chandler, P., Tuovinen, J. E., and Sweller, J. (2001). When problem solving is superior to studying worked examples. *J. Educ. Psychol.*, 93, 579–588.

Kalyuga, S., Ayres, P., Chandler, P., and Sweller, J. (2003). Expertise reversal effect. *Educ. Psychol.*, 38, 23–31.

Kirschner, P., Sweller, J., and Clark, R. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry-based teaching. *Educ. Psychol.*, 41, 75–86.

Klahr, D. and Nigam, M. (2004). The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning. *Psychol. Sci.*, 15, 661–667.

Leahy, W. and Sweller, J. (2005). Interactions among the imagination, expertise reversal and element interactivity effects. *J. Exp. Psychol. Appl.*, 11, 266–276.*

Low, R. and Sweller, J. (2005). The modality principle. In *Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer, pp. 147–158. Cambridge, U.K.: Cambridge University Press.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *Am. Psychol.*, 59, 14–19.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In *Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer, pp. 31–48. Cambridge, U.K.: Cambridge University Press.*

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 63, 81–97.

Paas, F. and van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: a cognitive-load approach. *J. Educ. Psychol.*, 86, 122–133.

Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educ. Psychol.*, 38, 1–4.*

Paas, F., Renkl, A., and Sweller, J. (2004). Cognitive load theory: instructional implications of the interaction between information structures and cognitive architecture. *Instruct. Sci.*, 32, 1–8.

Pollock, E., Chandler, P., and Sweller, J. (2002). Assimilating complex information. *Learn. Instruct.*, 12, 61–86.

Renkl, A. and Atkinson, R. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: a cognitive load perspective. *Educ. Psychol.*, 38, 15–22.

Renkl, A., Atkinson, R., and Große, C. (2004). How fading worked solution steps works: a cognitive load perspective. *Instruct. Sci.*, 32, 59–82.

Simon, H. A. and Gilmartin, K. (1973). A simulation of memory for chess positions. *Cogn. Psychol.*, 5, 29–46.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cogn. Sci.*, 12, 257–285.*

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learn. Instruct.*, 4, 295–312.*

Sweller, J. (2003). Evolution of human cognitive architecture. In *The Psychology of Learning and Motivation*, Vol. 43, edited by B. Ross, pp. 215–266. San Diego, CA: Academic Press.*

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instruct. Sci.*, 32, 9–31.

Sweller, J. (2005a). Implications of cognitive load theory for multimedia learning. In *Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer, pp. 19–30. Cambridge, U.K.: Cambridge University Press.

Sweller, J. (2005b). The redundancy principle. In *Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer, pp. 159–167. Cambridge, U.K.: Cambridge University Press.

Sweller, J. and Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cogn. Instruct.*, 2, 59–89.

Sweller, J. and Sweller, S. (2006). Natural information processing systems. *Evol. Psychol.*, 4, 434–458.

Sweller, J., Chandler, P., Tierney, P., and Cooper, M. (1990). Cognitive load and selective attention as factors in the structuring of technical material. *J. Exp. Psychol. Gen.*, 119, 176–192.

Tettamanti, M., Buccino, G., Saccuman, M., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S., and Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.*, 17, 273–281.

Tindall-Ford, S., Chandler, P., and Sweller, J. (1997). When two sensory modes are better than one. *J. Exp. Psychol. Appl.*, 3, 257–287.

Van Merriënboer, J. J. G. (1997). *Training Complex Cognitive Skills*. Englewood Cliffs, NJ: Educational Technology Publications.*

Van Merriënboer, J. J. G. and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.*, 17, 147–177.*

Van Merriënboer, J. J. G., Schuurman, J., de Croock, M., and Paas, F. (2002). Redirecting learner's attention during training: effects on cognitive load, transfer test performance and training efficiency. *Learn. Instruct.*, 12, 11–37.

* Indicates a core reference.