CHAPTER TWO

Cognitive Load Theory

John Sweller

Contents

Abstract

Cognitive load theory uses evolutionary theory to consider human cognitive architecture and uses that architecture to devise novel, instructional procedures. The theory assumes that knowledge can be divided into biologically primary knowledge that we have evolved to acquire and biologically secondary knowledge that is important for cultural reasons. Secondary knowledge, unlike primary knowledge, is the subject of instruction. It is processed in a manner that is analogous to the manner in which biological evolution processes information. When dealing with secondary knowledge, human cognition requires a very large information store, the contents of which are acquired largely by obtaining information from other information stores. Novel information is generated by a random generate and test procedure with only very limited amounts of novel information able to be processed at any given time. In contrast, very large amounts of organized information stored in the information store can be processed in order to generate complex action. This architecture has been used to generate instructional procedures, summarized in this chapter.

Psychology of Learning and Motivation, Volume 55

ISSN 0079-7421, DOI 10.1016/B978-0-12-387691-1.X0001-4

1. Introduction

Cognitive load theory is an instructional theory based on our knowledge of human cognition (Sweller, Ayres & Kalyuga, 2011). Since its inception in the 1980 s (e.g., Sweller, 1988), the theory has used aspects of human cognitive architecture to generate experimental, instructional effects. These effects are demonstrated when novel instructional procedures are compared with more traditional procedures as part of a randomized, controlled experiment. If the novel procedure facilitates learning, based on test performance, a new effect may have been demonstrated, an effect generated by our knowledge of human cognition. The new instructional procedures that follow from the effect become candidates for relevant professionals such as instructional designers and teachers.

While cognitive load theory is not unique in using human cognition to generate instructional procedures, it is regrettably rare for instructional design to be based on human cognitive architecture. Frequently, instructional design principles are promulgated as though human cognition either does not exist or if it does exist, it has no implications for instruction. An alternative to a theory-free process is to determine instructional design by using well-known cognitive structures such as working memory and long-term memory. These structures and their properties have strong implications for instruction. They can generate hypotheses that can be tested experimentally and if supported, can lead to new effects and novel instructional procedures.

Cognitive load theory, by using our knowledge of the relations between working memory and long-term memory, has been able to generate instructional procedures that to some can appear counterintuitive. Furthermore, a large range of those instructional procedures that otherwise would appear random and unconnected to each other can be seen to be closely related by their common, theoretical base provided by human cognitive architecture. That architecture, discussed in Section 2, not only indicates the relations between the instructional effects discussed in Section 3 but also provides an explanation why an effect is obtained and the conditions under which it can or cannot be obtained.

None of the experimental effects and the instructional procedures that flow from these effects is universal in the sense that it can be obtained under all conditions. All effects depend on variations in cognitive load. For this reason, the effects should not be considered in isolation from human cognitive architecture. An effect that occurs under one set of conditions may disappear under conditions that on the surface appear very similar but in fact differ substantially when considered from the perspective of the cognitive load imposed. Analyzing the instructional conditions discussed in Section 3 using the cognitive architecture of

Section 2 can explain why apparently similar conditions impose a differential cognitive load. As will be seen in Section 3, this analysis has frequently given rise to new experimental effects and so new instructional procedures. The next section will discuss those aspects of human cognitive architecture that have been incorporated into the theory.

2. Human Cognitive Architecture

In the last few years, cognitive load theory has taken an evolutionary view of human cognitive architecture (Sweller, 2003; Sweller & Sweller, 2006). There are two aspects of this treatment. First, the theory has incorporated Geary's (2007, 2008) categorization of knowledge into biologically primary and secondary knowledge. This categorization assumes that we have specifically evolved to acquire some particular types of information, known as biologically primary knowledge, while we have only needed other types of information, known as biologically secondary knowledge, in more recent times and so have not evolved a specific disposition to acquire that information. Only biologically secondary knowledge is the subject of instruction. Second, the theory has suggested that biologically secondary knowledge is acquired, organized, and in general processed in the same way as evolution by natural selection "processes" information (Sweller, 2003; Sweller & Sweller, 2006). Evolution by natural selection is normally and appropriately considered as a biological theory. In this chapter, it will be suggested that it should also be considered as a natural information processing system. Geary's categorization of knowledge according to its evolutionary status will be discussed next.

2.1. The evolutionary status of knowledge

Knowledge and skill can be classified into an enormous variety of categories. The vast majority of potential classification schemes have failed to yield instructional consequences in that instructional procedures that facilitate learning in one category may equally facilitate learning in another. In contrast, Geary's (2007, 2008) classification of knowledge into biologically primary and secondary knowledge is directly relevant to instructional procedures.

2.1.1. Biologically Primary Knowledge

Consider a young child learning to speak and listen to his or her native language. The child may be given considerable assistance by parents and others. They may repeat key words, speak clearly and distinctly, and use a very restricted range of vocabulary and grammar that heavily emphasizes "baby-talk." Nevertheless, children are not explicitly taught how to listen and speak. Indeed, with the exception of speech therapists, most people are likely to have little idea how to teach children to speak their native language. For example, children acquire the immensely complex motor actions associated with speech with no tuition whatsoever. An appropriate coordination of tongue, lips, breath, and voice occur without any explicit instruction. For most of us, simple membership of a functioning society is sufficient to learn to speak our native language. Despite the complexity of the task, we do not require explicit tuition.

Learning to listen and speak are biologically primary skills (Geary, 2007, 2008). They are skills that we have evolved to acquire over countless generations. We do not need to be motivated by others to acquire these language skills. We are self-motivated and acquire the skills easily, effortlessly, and unconsciously without instruction. We will automatically take on the accent of our society rather than the accent of, for example, immigrant parents because we have evolved to learn to speak with the accent of our peers.

There are many biologically primary skills. We learn basic social relations and we learn to recognize faces just as easily and automatically as we learn our native language. In each case, external motivational factors are irrelevant because we have evolved to acquire these skills and explicit instruction is unnecessary.

Biologically primary skills are modular. Learning our native language and learning to recognize faces require quite different, unrelated processes. We may have deficits in one biologically primary area with no apparent deficiencies in another. We may have evolved to acquire different biologically primary skills at very different times in our evolutionary history.

2.1.2. Biologically Secondary Knowledge

The nature and acquisition processes of biologically secondary skills are quite different from the processes associated with primary skills. We have evolved to acquire secondary skills but only in a general sense, not as specific modular abilities. Biologically secondary knowledge is knowledge that has become culturally important and needs to be acquired in order to function appropriately in a society. While listening and speaking provided examples of biologically primary knowledge, reading and writing provide equivalent examples of biologically secondary knowledge and can be used to demonstrate some of the characteristics of secondary knowledge.

As indicated above, most of us will learn to listen and speak simply as a consequence of living in a normal, listening/speaking society. In contrast, simply living in a reading/writing society is insufficient to allow most people to learn to read and write. Reading and writing became near universal skills only in some societies with the rise of modern education. The fact that a few people in some cultures could read and write was not sufficient to allow most people to read and write, a state of affairs that persisted for several thousand years. People will learn to listen and speak without explicit tuition. They will rarely learn to read and write without specific tuition.

The difference between listening/speaking and reading/writing is evolutionary. We have evolved to learn to listen and speak. We are able to learn to read and write, but we have not specifically evolved to read and write. The evolved perceptual motor and cognitive skills we use to read and write did not evolve in relation to reading and writing. The skills evolved for other reasons, but we are able to use these skills to learn to read and write. The vastly different evolutionary history of speaking/listening and reading/writing has both cognitive and educational consequences.

2.1.3. Consequences of the Distinction Between Biologically Primary and Secondary Knowledge

With respect to cognitive consequences, while we are internally motivated to learn to listen and speak, and learn to do so relatively effortlessly unconsciously and without external encouragement or explicit tuition, the same ease of acquisition is not apparent in the case of learning to read and write. We may not be motivated to learn to read and write and so learning reading and writing is likely to require considerable conscious effort over long periods of time. A considerable minority of people in a reading/writing culture may never learn to read and write.

The educational consequences of learning to read and write compared to learning to speak and listen are stark. If a society wants most of its people to read and write, it must specifically organize itself through its education systems to ensure that most of its members learn to read and write. We do not need educational systems and procedures to teach people to listen and speak. In contrast, without schools, most people will not learn to read and write. Schools and other educational and training institutions have been established to deal with biologically secondary knowledge such as reading and writing. Every subject taught in educational and training institutions can be virtually classified as incorporating biologically secondary knowledge.

There are other instructional consequences of the distinction between biologically primary and secondary knowledge. In recent years, there has been a heavy emphasis in the research literature on teaching general cognitive and metacognitive strategies. The distinction between primary and secondary knowledge casts some doubt on the relevance of that emphasis. First, it is difficult to find any cognitive or metacognitive strategies around which there is a consensus, rendering it difficult to assess the validity and usefulness of a particular strategy. Second, the whole point of cognitive or metacognitive strategies is that they are very general, applying to a vast array of tasks. It is much easier to classify them as biologically primary than secondary simply because we are likely to have evolved to acquire a skill that has very wide applicability. A metacognitive skill such as learning to organize information is likely to be essential for the survival of the human species. If a university student, for example, cannot organize information, it is more likely that he or she suffers from the complexity of the particular information with which he or she is dealing with rather than an ignorance of how to organize information. A biologically primary skill such as organizing information may not be teachable or learnable because it will have been already acquired by normally functioning people.

Of course, the above argument would be rendered irrelevant if we had a body of evidence based on randomized controlled experiments demonstrating the advantages of being taught general skills. This body of evidence is missing despite there being many studies demonstrating relatively improved performance following instruction in general skills. Unfortunately, almost without exception, these studies are flawed either because they alter multiple variables simultaneously and so eliminate any possibility of determining causality or because they fail to use far transfer tests. Transfer is essential if we are to exclude the possibility that the acquisition of domain-specific knowledge provides the factor determining improved performance. If, for example, we claim that learners who are taught how to organize knowledge subsequently will learn better than students who have not been taught to organize knowledge, we need to demonstrate that improved learning in areas quite unrelated to the area used to teach the strategy. The use of the same or a similar area in teaching and testing cannot exclude the possibility that learners have merely acquired domain-specific knowledge, rather than knowledge of how to organize information. To my knowledge, there is no scientifically acceptable body of evidence for any general strategy indicating that it is teachable and beneficial in a variety of unrelated areas. A likely reason for that failure may be found in the suggestion that general cognitive strategies consist of biologically primary knowledge.

2.2. The acquisition and organization of secondary knowledge

The manner in which the human cognitive system is organized to acquire, retain, and disseminate biologically secondary information is directly relevant to instructional design. As an example that has generated some controversy in recent years, if human cognitive architecture is better suited to discovering secondary information than receiving the same information, then instruction needs to be organized in a manner that encourages discovery. Alternatively, if the human cognitive system is better at acquiring information from other humans than discovering the same information, then instructional systems need to reflect that fact by emphasizing the presentation rather than the discovery of information (Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004). Furthermore, if we acquire biologically primary information in a manner that is very different from the manner in which we acquire biologically secondary information, then the distinction between the two categories of information becomes an important consideration. Accordingly, how we deal with information, especially biologically secondary information, is critical to instructional design.

There are many ways of approaching the issue of how the human cognitive system deals with information, with the most common being to study the components of human cognitive architecture such as working memory or long-term memory. Much of our knowledge about human cognition comes from such critically important work. Nevertheless, there is an alternative, complementary approach. Humans are, of course, part of nature and nature processes information. For example, while well-known theories such as evolution by natural selection are characteristically considered as biological theories, they can just as easily be considered as natural information processing theories (Sweller & Sweller, 2006).

Evolution by natural selection creates novel information, stores that information for later use, and disseminates it across space and time. It can be considered as an example of a natural information processing system. Biological evolution gave rise to humans including the human cognitive system. Unsurprisingly, given its evolutionary origins, the human cognitive system is also a natural information processing system with characteristics similar to that of the evolutionary system. When dealing with biologically secondary information, the human cognitive system also creates novel information, stores it for later use, and disseminates it across space and time.

The characteristics of natural information processing systems such as biological evolution and human cognition can be specified in a variety of ways. In this chapter, five basic principles will be used to describe the systems (Table 1). The *information store principle* indicates the role of stored information in the functioning of natural information processing systems, with the *borrowing and reorganizing principle* providing the major process by which information is acquired. The *randomness as genesis principle* indicates the centrality of random generate and test procedure to the creation of novel information. The importance of processing very small amounts of information when engaging in random generation is covered by the *narrow limits of change principle*, while the ability to handle very large amounts of previously organized information is dealt with by the *environmental organizing and linking principle.* Together, these principles indicate how information can be created, stored, disseminated, and used by natural information processing systems. Each of the principles will be discussed next.

Table 1 Natural Information Processing Principles

Principle	Function
Information store principle	Store information
Borrowing and reorganizing principle	Obtain information from others
Randomness as genesis principle	Generate novel information
Narrow limits of change principle	Restrict the random generation of novel information to protect the information store
Environmental organizing and linking principle	Use stored information to determine appropriate action within an environment

2.2.1. Information Store Principle

In order to function in the normally very complex natural environment, natural information processing systems require very large stores of information that can be used to direct appropriate activity. A genome provides that information store in the case of evolution by natural selection. While there is no agreed upon technique for measuring the size of a genome (Portin, 2002; Stotz & Griffiths, 2004), all genomes, even relatively simple ones, require thousands of units of information. More complex genomes may require billions of units of information. It is appropriate to consider a genome to be a very large information store designed to appropriately organize complex activity using complex processes.

The human cognitive system must also navigate a complex environment, and similar to evolution by natural selection, human cognitive architecture requires a large information store in order to function. Human long-term memory provides that store.

The central importance of long-term memory to cognitive functioning is often overlooked, especially in education. Long-term memory tends to be tacitly dismissed as consisting of little more than isolated, random elements of information. It is, of course, far more important than that depiction. Long-term memory is central to all cognitive functioning. The importance of long-term memory to general cognition became apparent only following seminal work in the field of problem solving. The work not only changed our view of long-term memory but also our view of problem solving and, indeed, the nature of human cognition.

The initial work was published by de Groot (1965) in the 1940 s, but it attracted a wider audience only when it was translated from the original Dutch to English and republished over 20 years later. De Groot was interested in the cognitive factors that distinguished chess grandmasters from weekend players. We know chess grandmasters virtually always defeat weekend players, but it was not clear what skills they had developed to allow this superiority. Better problem-solving skill provided the most likely hypothesis, but we did not really know what that meant. De Groot investigated some obvious hypotheses. Chess experts may plan ahead a larger number of moves than weekend players. That is, they may engage in a greater search in depth. Alternatively, they may consider a greater number of possible moves at each choice point, indicating a greater search in breadth. In the case of either those who increase their search in depth or in breadth, we might expect them to increase their chances of finding good moves and hence to increase their probability of winning. De Groot tested the hypotheses that expert chess players engage in a greater search in depth or breadth compared to weekend players, but essentially found no differences on these measures between different grades of players. Whatever cognitive processes the chess grandmasters engage in to win, looking further ahead or considering a greater range of possible moves than the weekend players was not included in their repertoire of skills.

There was one difference between chess grandmasters and weekend players that De Groot found. He presented grandmasters with a board on which the pieces had been placed in an arrangement taken from a real game. In other words, the board configuration was one that could be found during a game. De Groot showed the grandmasters the board configuration for 5 s before taking it away and asked them to replicate the configuration that they had just seen. They were surprisingly good at this task, accurately replacing about 70% of the pieces. In contrast, weekend players were much poorer, accurately replacing only about 30% of the pieces. Chase and Simon (1973) replicated these findings and found that if the pieces were placed on the board in a random configuration, the differences between grandmasters and weekend players disappeared. Both were able to accurately replace only about 30% of random board configurations. The superiority of chess grandmasters was restricted to board configurations taken from real games.

What do these results tell us about skill in chess in the first instance and, more generally, about long-term memory, problem solving, and cognition? It takes at least a decade to become a chess grandmaster (Ericsson & Charness, 1994). During this period, grandmasters not only play many games but also spend many hours each day studying previous games. While studying and playing games, grandmasters learn to recognize a large number of board configurations and the best moves associated with each configuration. Chess is a game of problem solving, but chess grandmasters' skill does not derive from some mysterious and undefinable problem-solving skill. Rather it derives from a familiarity with a great number of board configurations and the moves associated with those configurations. A chess grandmaster does not have to plan a sequence of moves because he or she knows which moves work well and which do not. It is a weekend player who must plan moves because he or she does not have the large repertoire of moves acquired by grandmasters. The repertoire of moves is held in long-term memory and explains the skill of chess grandmasters. No other skill, particularly no general problem-solving strategies, has been found to differentiate chess grandmasters and weekend players. Neither are other skills required to explain the performance of chess grandmasters.

Grandmasters have been estimated to hold tens of thousands of board configurations in long-term memory (Simon & Gilmartin, 1973). Although impressive, we need to note that many educated people have similar skills due to similar stores of information held in long-term memory but in areas other than chess. The skill exhibited by chess grandmasters is unusual because few people become professional chess players. In contrast, many more people become competent mathematicians and even more learn to read competently. If de Groot or Chase and Simon had demonstrated that competent readers can readily reproduce the letters "chess is a game of problem solving skill" while poor readers or non-readers are far poorer, the influence of the finding would likely have been much reduced. Similarly, demonstrating that the letters "lliks gnivlos melborp fo emag a si ssehc" are equally poorly remembered by good and poor readers is less likely to attract attention than the same finding from the game of chess. The genius of de Groot's and Chase and Simon's findings was their demonstration of the critical importance of long-term memory in an area in which the influence of long-term memory was assumed to be negligible.

The results from chess have been extended to a variety of educationally relevant areas such as understanding and remembering text, designing software, and solving mathematical problems (Chiesi, Spilich, & Voss, 1979; Egan & Schwartz, 1979; Jeffries, Turner, Polson, & Atwood, 1981; J. Sweller & Cooper, 1985). Experts in a given area are able to better remember information associated with that area and are able to better use that information to solve problems. They recognize problem states and the best moves associated with each state.

There are consequences associated with these findings both for our understanding of human cognition and for educational research and practice. With respect to human cognitive architecture, the role of long-term memory is transformed. We do not use long-term memory just to remember items. We use it to determine the bulk of our activity. If we are good at something, it is because we have stored innumerable elements of information concerning that area in our long-term memory. All expertise, on this view, is determined by what is stored in the long-term memory. Activities such as problem solving that traditionally were assumed to be largely unrelated to the characteristics of long-term memory can now be seen to closely depend on it. Long-term memory is a central to, perhaps *the* central structure in, human cognitive architecture.

The nature of learning is also changed by this perspective. A competent person is not someone who has acquired complex, sophisticated, cognitive strategies that can be used in a variety of unrelated areas. Such teachable/learnable general strategies have not been described, probably because they are biologically primary. Rather, competence is domain specific. Some domains may be applicable in a large variety of areas, but it is still the particular domain that is important. Acquiring chess skill may not be usable in any areas other than playing chess. Acquiring reading skill may allow a person to read an unlimited number of unrelated texts. Nevertheless, reading skill applies only to reading. It will not improve a person's chess skill. Neither, in isolation, will it improve a person's knowledge of history. It will, of course, enable one to read historical texts and being able to read historical texts will improve a person's knowledge of history. The point is that in each case, the skill can be clearly specified and so clearly taught. From an educational perspective, the role of education is to increase knowledge held in long-term memory of particular discipline areas. How that knowledge is best acquired is a concern of cognitive load theory. Teaching general cognitive skills, on the other hand, may need to await the specification of such skills.

2.2.2. Borrowing and Reorganizing Principle

If, as indicated by the information store principle, natural information processing systems require a very large store of information in order to function in complex natural environments, the processes by which large stores of information are acquired become a critical issue. In the case of evolution by natural selection, the processes of reproduction, both asexual and sexual, are well known. They constitute the primary procedures by which a store of information is acquired. During asexual reproduction, each genome is copied exactly from the genome of the previous generation, with the exception of occasional mutations. In this sense, the information store that constitutes a particular genome has been borrowed largely in its entirety from the information store of the preceding generation. Borrowing can be seen to be a major procedure for acquiring a large information store.

Borrowing is equally important in the case of sexual reproduction. The major difference between asexual and sexual reproduction is that in the case of asexual reproduction, an information store is borrowed with no or minimal change, while in the case of sexual reproduction, the information is reorganized. Not only is the information reorganized, in addition, that reorganization provides the major reason for the existence of sexual reproduction. The reorganization of information during sexual reproduction results in a logical structure ensuring that each generation is necessarily different from the previous generation. Sexual reproduction occurs in order to ensure that, unlike the case of asexual reproduction, generational variation is a logical necessity of the procedure. During sexual reproduction, information is obtained and combined from both male and female parents resulting in offspring that necessarily differ from either parent. Information is not only borrowed, it is also reorganized.

The acquisition and storage of information in long-term memory more closely resembles sexual than asexual reproduction. We rarely remember information with minimal or no change in the same way as asexual reproduction or as an electronic recording device "remember" information. We do acquire or borrow the vast bulk of the information held in long-term memory from other people, but we alter that information depending on what we have already stored in long-term memory.

The processes by which we borrow information from others are well known. We imitate other people (Bandura, 1986), listen to what others tell us, read what they write, and look at diagrams and pictures that they produce. Listening, reading, and looking at diagrams and pictures are particularly important in the acquisition of the biologically secondary information that is the subject of education and training. The vast bulk of the biologically secondary information that is stored in long-term memory is acquired by one of the processes or a combination of these processes.

Although information is borrowed from others, it is reorganized in a manner analogous to sexual reproduction. We combine new information with information already stored in long-term memory and it is the new, reorganized information that is stored rather than an exact copy of the information that was presented. In other words, we store information as schemas rather than as precise copies (Chi, Glaser, & Rees, 1982). Each schema stored is likely to be different from the schema held in the long-term memory of the person from whom it was borrowed because it is a combination of the borrowed information combined with information already held in long-term memory.

Cognitive load theory has been used to generate many instructional effects and these effects rely heavily on the borrowing and reorganizing principle. The effects are largely concerned with techniques for presenting information to learners that are most likely to result in the facilitation of schema acquisition. In addition to the acquisition of schemas, cognitive load theory is also concerned with their automation so that they can be used without conscious processing in working memory (Kotovsky, Hayes, & Simon, 1985; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). In the case of schema acquisition, the theory assumes that learners acquire domain-specific information that is best obtained from other people. All the cognitive load instructional effects depend on these assumptions.

The theory has been concerned with the generation of novel information, the topic of the next section, to a far lesser extent. There are good reasons for that emphasis on obtaining information from others rather than generating it oneself. Contrary to many educational assumptions, humans rarely generate novel information (Kirschner et al., 2006). We almost always prefer to obtain information from others if it is available. If information is required but not available from others, only then do we need to generate it ourselves by, for example, conducting research.

In contrast, constructivist teaching procedures place a far heavier emphasis on learners generating information. The emphasis on generating information during constructivist learning and teaching in the past two decades ignores much of what we have learned about human cognition. There are two issues: whether we need to be taught how to construct knowledge and whether knowledge we have constructed during constructivist teaching sessions is superior to knowledge we have acquired from others.

With respect to teaching learners how to construct knowledge, while we must construct schematic knowledge in long-term memory in order to learn, it is a fallacy to assume that we need to be taught how to construct knowledge. We have evolved to construct knowledge. It is a biologically primary skill. There is no body of evidence based on properly conducted, randomized, controlled studies that teaching learners how to construct knowledge results in better learners. Neither is there a body of experimental evidence that teaching procedures such as discovery learning that require learners to discover or generate information for themselves constitute a better form of learning in comparison to the explicit presentation of information. Evidence that information that is discovered for ourselves is superior to studying the same information presented to us is missing (Klahr & Nigam, 2004). We have neither theoretical reasons nor empirical evidence that withholding information from learners results in better learning. On the contrary, based on the worked example effect, discussed below, we have strong evidence from a large variety of learning areas using very young to adult students that learning is facilitated by direct, explicit instruction.

2.2.3. Randomness as Genesis Principle

As indicated by the borrowing and reorganizing principle, natural information processing systems have powerful techniques for disseminating information, but that information must be created in the first instance in order to have something to disseminate. The randomness as genesis principle provides the necessary machinery for creating novel information.

In biological evolution, random mutation is the ultimate source of all genetic variation. While there are a variety of mechanisms, such as sexual reproduction, for handling and distributing the variations that occur due to mutation, without mutation none of these mechanisms could function. For example, sexual reproduction relies on combining different alleles from male and female parents. Random mutation accounts for the fact that alleles differ. If they did not differ, combining them would have no function. Ultimately, evolution by natural selection assumes that all the variation, not only within species but also between species, can be sourced to random mutation.

There are important consequences that flow from the fact that all genetic variation ultimately derives from random mutation. Evolution by natural selection is a creative system. It has created the entire biological world. The source of that creativity is the randomness as genesis principle. The randomness as genesis principle has a basic problem-solving process, random generate and test, as its creative engine. Random generation creates novelty and it is this novelty that has given rise to the immense diversity of the biological world. Nevertheless, the "test" part of random generate and test is just as important to creativity as the "random generation" part. While random mutation is essential, in isolation it would not and could not generate the diversity and complexity that we see in biological structures and functions. Mutations are randomly generated, but whether a mutation has any substantive biological consequences depends on whether it is adaptive. If a mutation increases the adaptivity of an organism to its environment, it is likely to be retained for future generations. In other words, it is added to the information store, in this case a genome. If it is maladaptive, it is not added to the information store and is likely to be lost. In this manner, mutations are tested for effectiveness with effective mutations added to the genome and retained while ineffective mutations are jettisoned. Thus, when applied to evolution by natural selection, the randomness as genesis principle is closely tied to a problem-solving process, random generate and test.

The randomness as genesis principle functions in an analogous way in human cognition and is equally important (Sweller, 2009). While most of the knowledge held in long-term memory is acquired via the borrowing and organizing principle, the knowledge is created in the first instance during problem solving. When dealing with familiar problems, problem solving largely consists of retrieving schematic information from long-term memory. Our schemas allow us to recognize a problem as belonging to a particular class of problems that require a particular solution (Chi et al., 1982). Dealing with familiar problems in this manner is critical to problem-solving skill but is unlikely to result in the generation of new knowledge. In contrast, dealing with novel, unfamiliar problems has the potential to create new knowledge. New knowledge can be generated when we discover a new procedure or concept during problem solving. Random generate and test is central to solving unfamiliar problems. All problem-solving procedures intended to deal with novel problems, at some point, incorporate a random generate and test process that is indistinguishable from the random generate and test process used by evolution by natural selection.

The logical status of random generate and test needs to be considered. It is argued that when faced with a potential problem-solving step that we have not previously carried out, or faced with an entire problem for which we do not have a solution stored in long-term memory, then we have no choice but to engage in a random generate and test procedure. Assume that knowledge is unavailable to generate a known move or assume that there is sufficient knowledge to generate two or more potential moves but insufficient knowledge to rank them in terms of their likelihood of success. Faced with the lack of information held in long-term memory, we must randomly choose a move and attempt to test that move for effectiveness. There appear to be no logical techniques available to generate a move under conditions where knowledge is unavailable other than random generation.

We can rationally deduce moves, but all techniques for doing so require knowledge. As an example, we may have a new algebra problem that we have not seen previously but which conforms to the structure $a/b = c$, solve for a. Our knowledge, stored in long-term memory, tells us that the new problem conforms with the structure of $a/b = c$, solve for a, and we also know how to solve problems of this type. We can use this knowledge to generate a problem solution.

Without this knowledge, generating a solution would be more difficult and random generate and test provides the only available generative technique. We can use the rules of algebra to try several moves until we find one that either solves the problem or takes us closer to solving the problem. We might attempt to subtract b from both sides or add b to both sides, but discover these moves are ineffective. We might then discover that multiplying both sides by b is effective and solves the problem.

It should be noted, that even with knowledge, it might be argued that aspects of random generate and test are being used. Assume that we are solving the problem analogically. If we have not seen the relevant problem previously, we cannot be certain that it really does conform to the structure of $a/b = c$, solve for a. In other words, there can be no certainty that the analogy works. We can only be certain that the two problems are analogous once we have chosen the move of multiplying out the denominator and checked to see if it works. Certainty is impossible prior to making the relevant move or moves. Often we may only find that the problem looks as though it is a problem of a certain type but when we try to solve it accordingly, we may discover that the solution does not work, an example of Einstellung (Luchins, 1942; Sweller & Gee, 1978). Einstellung occurs when problem solvers, categorizing a problem incorrectly, fail to see a very simple solution and attempt to solve the problem using a complex or, in extreme cases, an impossible solution. Even when using knowledge to generate a successful solution, we can have no certainty that the knowledge is relevant and that it is properly used until we have attempted to use it.

In the absence of appropriate knowledge, the knowledge may need to be generated and random generate and test is the only available procedure. If alternatives to random generate and test when faced with novel problems are suggested, the procedures must be specified. To this point, random generate and test is the only generative process that has been specified when faced with novel problems for which a complete series of moves is not available in long-term memory.

The randomness as genesis principle provides the source by which new knowledge is created (Sweller, 2009). Once created and shown to be effective, that knowledge can be stored in the information store. While likely to be quantitatively, comparatively small, it has the same status as knowledge stored via the borrowing and reorganizing principle. From evolution by natural selection, we know it is the ultimate source of creativity in the biological world and we also know by observing the biological world that it is a highly effective source of creativity.

There is every sign that the randomness as genesis principle plays the same role in human cognition as in evolution by natural selection. If so, there are educational implications. Calls to encourage generative processes in education or to encourage creativity need to be made in light of the nature of generative processes and creativity. We need to understand that teaching learners to be flexible and creative requires us to teach them to engage in random generate and test. At this point, it is unclear whether encouraging learners to engage in random generate and test is likely to be productive. The question needs to be answered using appropriate experiments. Simply asserting that encouraging learners to engage in generative, constructivist, creative activities will be beneficial is inappropriate in the absence of data.

2.2.4. Narrow Limits of Change Principle

The randomness as genesis principle has structural implications. Random generation and test is concerned, in all cases, with the manner in which elements of information should be combined. Some combinations of elements prove to be effective when tested, others do not. The number of combinations that need to be tested can be critical. For example, there are six permutations of three elements ($3! = 6$). In contrast, there are 3,628,800 permutations of 10 elements ($10! = 3{,}628{,}800$). A random generate and test process that must find an appropriate permutation of 10 elements is vastly more difficult than a random generate and test process that must find the permutations of 3 elements. The implication of this arithmetic is that random generation and test should only deal with very small numbers of elements at a time.

This logic is directly relevant to the randomness as genesis principle and so structures are required that take that logic into account. The narrow limits of change principle provides these structures. The randomness as genesis principle is concerned with how natural information processing systems deal with novel information not previously stored in the information store. The novel information is obtained from the external environment and so the structure provided by the narrow limits of change principle is needed to deal with information from the external environment. In the case of evolution by natural selection, the relevant structure is the epigenetic system (Jablonka & Lamb, 2005; West-Eberhard, 2003). This system intercedes between the genetic system and the external environment. It manages the interaction between the genetic system based on DNA and the environment external to the DNA. The epigenetic system may be as equally important as the genetic system, although much less is known of it than the genetic system. Both systems are distinct (Jablonka & Lamb, 1995, 2005; West-Eberhard, 2003) and although they act independently, they closely interact.

The epigenetic system is able to transmit information from the external environment to the DNA-based genetic system in order to affect genetic alterations. Information from the environment can alter DNA by affecting when and where mutations occur. Environmental signals can facilitate or inhibit mutations in particular parts of a genome. For example, stressful environments may require changes in a genome in order to deal with the stress. These changes can occur via mutations and some organisms are able to increase the number of mutations when they find themselves in stressful environments. With increased mutations, there is a greater likelihood of a change to a genome that increases the chances of survival. As another example, mutations may be thousands of times higher than the average in some sections of a genome. Venomous animals such as snakes need to frequently change the composition of their venom to ensure their prey do not become immune to it. The epigenetic system can both facilitate these mutations and ensure they are not repaired.

It needs to be noted that while the epigenetic system can determine when and where mutations occur, it cannot determine the nature of a particular mutation. Beyond the epigenetic system's determining influence, each mutation is random and must be tested for effectiveness before being added to the DNA-based information store. Critically, even where the rate of mutations is increased, mutations are relatively rare. For the reasons outlined above, random generation and test must result in small changes to the genome. Large changes are likely to have catastrophic effects on the current store of information found in DNA because, based on the above arithmetic, there are a huge number of large changes that are possible and only a very few of these changes are likely to be adaptive. Accordingly, all effective changes are small and incremental.

The human cognitive system similarly must reduce the number of novel elements with which it deals for the same arithmetic reasons that apply to evolution by natural selection. In the case of human cognition, the relevant structure is human working memory. We probably know more about human working memory than about the epigenetic system because working memory has been studied more intensively for a longer period (Miller, 1956) than the epigenetic system. In particular, we have known for a long time that working memory, when dealing with novel information, is very limited in both capacity (Miller, 1956) and duration (Peterson & Peterson, 1959). These are exactly the limitations to be expected given the logic of dealing with novel information.

One of the major functions of working memory is to act as a conduit between the external environment and long-term memory in the same way as the epigenetic system acts as a conduit between the external environment and the DNA-based genetic system. The characteristics that we normally associate with working memory, its capacity and temporal limitations, occur when working memory must deal with novel information from the external environment. We know that working memory is unable to store more than about seven items of novel information (Miller, 1956) for more than about 20 s (Peterson & Peterson, 1959).

The processing capacity of working memory is considerably less than its storage capacity with no more than about three–four items of information being able to be processed at a time (Cowan, 2001). Processing refers to combining, contrasting, or dealing in some manner with multiple elements. The processing capacity limits of working memory are the limits we must expect of any natural processing system that must deal with novel information using a random generate and test procedure.

The narrow limits of change principle is critical to instruction and central to cognitive load theory. Instructional procedures need to take into consideration the capacity and duration limits of working memory. Recommended procedures that unnecessarily increase working memory load run the risk of severely constraining the ability of students to learn, where learning is defined as a positive change in long-term memory. Information that cannot be fully processed in working memory cannot be fully transferred to long-term memory inhibiting learning. Too many instructional recommendations proceed as though we do not have a working memory or if we do have a working memory, it is irrelevant to instructional considerations. At least in part, cognitive load theory was developed as an alternative to such instructional recommendations.

2.2.5. The Environmental Organizing and Linking Principle

The epigenetic system and working memory not only deal with novel information from the external environment but also use information from the external environment to organize information in the information store and determine how that information is to be used and translated into action. The characteristics of the epigenetic and working memory systems are vastly different when organizing novel information from the environment compared to when using environmental information to organize the information store. The environmental organizing and linking principle covers the relation between the external environment and the information store. It permits a natural information processing system to use environmental signals to determine appropriate action. This principle is the final, natural information processing principle and provides the ultimate justification for the preceding principles.

The importance of the epigenetic system in organizing the genetic system can be demonstrated readily. A major function of the epigenetic system is to turn genes on and off. Consider the genetic material that can be found in the nuclei of human cells. For a given person, the nucleus of each cell has exactly the same genetic material as the nucleus of every other cell for those cells that contain nuclei. For example, the nucleus of a skin cell has exactly the same DNA as the nucleus of a liver cell, barring mutations. Of course, the structure and function of a skin cell bears little resemblance to the structure and function of a liver cell. If the genetic structure of these two cells is identical, what causes the immense differences in their characteristics? The answer is the epigenetic system. This system, via the environment external to the nucleus that holds the genetic material, controls which genes are to be turned on and which genes are to be turned off. By selectively turning genes on and off depending on environmental signals, vastly different cell structures with vastly different functions are built despite all cells having an identical genetic structure. In this sense, the epigenetic system is at least as important in biological systems as the genetic system.

The epigenetic system, when influencing the rate or location of mutations, must deal with relatively small amounts of information at a time, for reasons indicated above when discussing the narrow limits of change principle. In contrast, when the epigenetic system deals with the previously stored and previously organized information of the genetic system, the strictures imposed by a random generate and test process are absent. Accordingly, there are no limits to the amount of genetic material that can be dealt with by the epigenetic system. Very large amounts of DNA that constitute some genes can be turned on or off by the epigenetic system.

It can be seen that the epigenetic system links environmental signals to the genetic system. In this sense, it links the environment to the information store. The environmental organizing and linking principle is the general principle used by natural information processing systems to allow signals from the environment to influence the operation of the information store. Working memory has the same role in human cognition as the epigenetic system has in biological systems. Working memory uses signals from the environment to determine which aspects of long-term memory are relevant to current processing. For example, assume that we are familiar with problems of the form $(a + b)/c = d$, solve for a. When we see a problem of this form, it acts as a signal or cue triggering those aspects of long-term memory relevant to this particular problem with the rest of long-term memory left unaffected. In this manner, working memory determines which aspects of long-term memory are triggered and which are ignored. Its function is identical to the epigenetic system in biological systems.

As is the case for the epigenetic system, the characteristics of working memory are very different when it is dealing with stored, previously organized information compared to when it is dealing with novel information from the environment. The capacity and duration limits of working memory found when it deals with novel information disappear when working memory deals with information from long-term memory. Just as there are no known limits to the amount of stored DNA that can be handled by the epigenetic system, there are similarly no known limits to working memory when it processes familiar information organized in a familiar manner, that is, information stored in long-term memory. In other words, there are no known limits to the amount of organized information held in long-term memory that can be cued by appropriate environmental signals.

The different characteristics of working memory when dealing with familiar as opposed to novel information has resulted in some theorists suggesting a different structure when working memory handles familiar as opposed to novel material. Ericsson and Kintsch (1995) suggested "long-term working memory" as the structure that accounts for the manner in which working memory handles previously learned information held in long-term memory. Long-term working memory describes the characteristics of working memory when it deals with information stored in long-term memory. Because these characteristics bear little resemblance to the characteristics of working memory when it deals with novel information from the environment, we must either postulate different structures to deal with familiar and unfamiliar information or postulate different processes engaged in by the same structure. With respect to current concerns, either characterization results in an identical outcome. Information held in long-term memory allows us to carry out actions that we otherwise could not possibly consider.

The environmental organizing and linking principle provides the ultimate justification for natural information processing systems. Via this principle, the information created by the randomness as genesis and narrow limits of change principles, transmitted by the borrowing and reorganizing principle and stored by the information store principle, can be used to determine action that is appropriate to a particular environment. This action provides the purpose for a natural information processing system.

Based on the conception of a natural information processing system, the purpose of instruction is to increase biologically secondary knowledge held in long-term memory. That knowledge changes us. It changes the characteristics of working memory by eliminating its capacity and duration limits and allows us to engage fluently and efficiently in actions that we otherwise could not dream of carrying out. Cognitive load theory uses this cognitive architecture to devise instructional procedures.

3. Element Interactivity and Categories of Cognitive Load

Biologically secondary information varies in the extent to which it imposes a working memory load. There are two basic sources of instructional cognitive load. Some information imposes a heavy cognitive load because of its intrinsic nature. That load is referred to as intrinsic cognitive load. It can only be changed by changing what is learned or by changing the knowledge levels of learners. Other information imposes a heavy cognitive load not because of its intrinsic nature but rather because of the way it is presented. That load is referred to as extraneous cognitive load. It can be reduced by changing the instructional procedures. Both categories of cognitive load are determined by the same underlying factor: element interactivity (Sweller, 2010). High element interactivity occurs when learners process a large number of elements of information simultaneously in working memory with low element interactivity requiring few elements. The number of elements of information being processed due to the intrinsic nature of the information determines intrinsic cognitive load, while the number of elements of information due to instructional design factors determines extraneous cognitive load. Details concerning intrinsic cognitive load will be discussed next.

3.1. Intrinsic cognitive load

Intrinsic cognitive load refers to the complexity of the knowledge that is being acquired without reference to how that knowledge is acquired. How knowledge is acquired refers to extraneous cognitive load and will be discussed below. One of the critical features of intrinsic cognitive load is that it is fixed and unalterable for given information to be processed by learners with given levels of expertise. Because intrinsic cognitive load refers to the intrinsic complexity of the information being processed, it cannot be altered other than by altering what is learned or the levels of expertise of the learners. Once knowledge that is to be learned and what the learner already knows are determined, intrinsic cognitive load is fixed.

We can determine levels of intrinsic cognitive load by determining element interactivity. Some information is very high in element interactivity and so imposes a very high working memory load, while other information is low. For example, consider students who must learn chemical symbols. There are many symbols and the task is difficult. Nevertheless, the difficulty is not caused by a heavy intrinsic cognitive load and so working memory is not overloaded by this task. Each symbol can be learned independent of every other symbol because there is minimal element interactivity between the learning elements. For example, students can learn that Cu is the symbol for copper without any reference to the fact that the symbol for iron is Fe. Working memory resources can be devoted entirely to learning the symbol for copper without any reference to other symbols. Element interactivity is low and so working memory load due to intrinsic cognitive load is also low.

In contrast, other information can be very high in element interactivity, imposing a high working memory load due to a high intrinsic cognitive load. As an example, learning to balance a chemical equation requires consideration of a large number of elements of information in working memory simultaneously. When dealing with any unfamiliar equation in any discipline area, element interactivity is likely to be high. No change can be made to any element of information in an equation without considering the consequences of that change for every other element in the equation. Since the elements interact, all elements must be considered simultaneously prior to any manipulation of an equation.

Consider students who must learn to solve the algebra problem, $(a + b)/c = d$, solve for *a*. In order to understand and solve this problem, each of the elements that constitute the problem must be processed in working memory. Because they interact, they cannot be processed serially. They must be processed simultaneously. Each algebraic symbol must be considered in relation to every other algebraic symbol and the problem goal. For novice algebra students, these interacting elements may overload working memory resulting in a failure to solve the problem. This heavy working memory load is not caused by the need to process many elements, but rather by the need to process many elements simultaneously. Some tasks, such as learning the chemical symbols, require many more elements to be processed and so are difficult. This difficulty has a cause different from that of the difficulty imposed by the need to process many elements simultaneously. Simultaneous processing imposes a heavy working memory load, while successive processing does not. Whether information can be processed simultaneously or successively depends on element interactivity.

Levels of expertise also determine element interactivity via the information store and environmental organizing and linking principles. For readers of this chapter for whom the above algebra problem may be familiar because they hold a schema for the equation in long-term memory, the problem and its solution may be processed with little working memory load. The load may be so low that the problem may be solved without recourse to written materials because the interacting elements are incorporated in a schema that can be treated as a single element in working memory. A schema, due to the environmental organizing and linking principle, allows us to readily remember the equation, the problem goal, and to correctly manipulate the equation in working memory because the schema held in long-term memory includes the original problem state and all subsequent states.

Cognitively, learning to balance a chemical equation or manipulate an algebraic equation is analogous to learning to make a good move in chess. In each case, there are many elements of information that must be processed simultaneously in working memory. If these elements are not incorporated into a schema that can be treated as a single element using the environmental organizing and linking principle, the element interactivity and intrinsic cognitive load will be high. Learning requires the acquisition of large numbers of schemas incorporating interacting elements and stored in long-term memory via the information store principle. Once stored, they can be transferred to working memory via the environmental organizing and linking principle, thus permitting cognitive activities that otherwise would be impossible to even contemplate.

3.1.1. Understanding

This analysis of element interactivity and intrinsic cognitive load can be used to explain understanding and the distinction between learning with understanding and learning by rote. Counterintuitively for some, long-term memory is central to understanding and this fact has bedevilled an analysis of the concept. *Understanding* does not apply to low element interactivity information. It applies exclusively to high element interactivity information. For example, with respect to low element interactivity, if a learner is unable to indicate the symbol for copper, we might say they have forgotten the symbol or never learned the symbol, but we would not refer to the failure as a failure of understanding. The role of memory is clear-cut and obvious in the case of low element interactivity material. In contrast, if a person is unable to balance a chemical equation or solve an algebra problem, the term *understanding* is readily applied. It is quite appropriate for us to refer to a person understanding or not understanding an equation. It is inappropriate to refer to understanding a chemical symbol. Nevertheless, the difference between knowing a correct symbol and knowing how to deal with an equation can be expressed entirely in element interactivity terms. The cognitive processes in both cases are identical with both relying on memory.

Consider a student learning to multiply two numbers such as $3 \times 4 = 12$. Some students may treat learning this process as nothing more than memorizing the answer to 3 × 4. If so, the task is treated as a low element interactivity or "rote-learned" task. Other students may learn that 3 × 4 means 3 lots of 4 or 4 + 4 + 4. These learners are beginning to *understand* the procedure. But note the process of understanding. It relies on long-term memory in exactly the same way as the rote learning with the only difference being in what is memorized. Rote learning simply means learning that 3 × 4 = 12, while learning with some degree of understanding means that in addition to learning that 3 × 4 = 12, students have also learned that 3 × 4 = 4 + 4 + 4 = 12. Both learning by rote and learning with understanding require changes to long-term memory with the only difference being that learning with understanding requires that more be memorized. If more is memorized, for example, that 3 × 4 = 4 + 4 + 4 = 3 + 3 + 3 + 3 = 12, then even more is understood. Further understanding occurs when 3 × 4 = 12 can be related to subtraction, division, and more general mathematical systems. In each case, further understanding consists of more information stored in long-term memory.

Before this high element interactivity information can be stored in long-term memory in order for the environmental organizing and linking principle to apply, it must of course be processed first in working memory. Processing high element interactivity information in working memory imposes a high intrinsic cognitive load. Acquiring the information requires a greater use of either the randomness as genesis principle if the information is discovered by learners or the borrowing and reorganizing principle if the information is presented. In either case, the working memory load (narrow limits of change principle) is increased compared to not having to process the additional information, especially if the randomness as genesis principle must be used. Learners can avoid processing the additional information by just learning that 3 × 4 = 12, resulting in a high element interactivity task being turned into a low element interactivity task. Of course, what has been learned has been changed. Changing what is learned from high to low element interactivity has the obvious advantage of reducing intrinsic cognitive load. There are obvious disadvantages to reducing intrinsic cognitive load when learning by rote instead of learning with understanding. Nevertheless, some students under some circumstances may have little choice in the matter. They may be unable to process the large number of interacting elements that need to be processed in order to learn with understanding. The intrinsic cognitive load imposed by learning with understanding may be overwhelming.

3.2. Intrinsic cognitive load effects

Cognitive load theory has been used to generate a large number of instructional procedures designed to alter cognitive load and, indeed, the generation of novel instructional procedures provides the ultimate purpose of the theory. A cognitive load effect is demonstrated when the theory is used to suggest ways of altering the number of interacting elements resulting in a new instructional procedure with better test outcomes than a traditional procedure. Most cognitive load effects are due to reductions in extraneous cognitive load (see below). There are few intrinsic cognitive load effects because intrinsic cognitive load cannot be altered except by altering the nature and goals of what is learned or by altering the levels of expertise. The variability effect and the isolated elements effect provide examples of effects due to changing levels of intrinsic cognitive load. Table 2 lists the cognitive load effects discussed in this chapter.

3.2.1. The Variability Effect

The variability effect, unlike all other cognitive load effects specified to date, occurs due to an increase rather than a decrease in cognitive load, in this case intrinsic cognitive load. Assume that learners are presented with a set of problems that are very similar. For example, they may vary only in the numerical values that need to be plugged into equations. In contrast, assume another set of problems in which, in addition to numerical values changing, equations have to be manipulated. The second set has greater variability resulting in increased element interactivity since more elements must be processed. Intrinsic cognitive load is increased because learners must not only learn how to solve a particular class of problems but must also learn to distinguish between problem types and learn which types require essentially the same solution and which types require a different solution. Providing that learners have sufficient working memory capacity to process the additional elements, there should be advantages to learning with more rather than less variable problems.

Paas and van Merrienboer (1994) obtained the variability effect with learners provided more variable problems learning more and performing better on transfer problems than the learners provided less variable problems. The effect is due to intrinsic cognitive load because what students were required to learn changed resulting in a change in element interactivity due to changed goals. Rather than just learning how to use an equation, a task that is relatively low in element interactivity, learners also had to learn which equations were appropriate at which time, a task that requires the processing of many more interacting elements. In terms of the cognitive architecture discussed in Section 2, increasing variability increased the amount of information stored so increasing the effectiveness of the environmental organizing and linking principle. The cost is an increased working memory load and so the procedure can be effective only if sufficient working memory resources are available.

Table 2 Cognitive Load Theory Effects

Effect	Description
Variability	Under low intrinsic cognitive load, increased variability increases intrinsic load resulting in increased learning if working memory resources are available
Isolated elements	Under high intrinsic cognitive load, presenting interacting elements as though they are isolated can decrease intrinsic load
Goal-free	Eliminating a problem goal eliminates the use of means–ends analysis reducing extraneous cognitive load
Worked example	Demonstrating a problem solution reduces the extraneous cognitive load associated with problem solving
Split-attention	If mental integration is required, extraneous cognitive load may be reduced by physically integrating disparate sources of information
Modality	Mental integration can be facilitated by presenting material using audiovisual rather than a visual only format
Redundancy	Processing unnecessary information imposes an extraneous cognitive load
Element interactivity	If intrinsic cognitive load is low, a high extraneous cognitive load may not exceed working memory capacity, reducing extraneous cognitive load effects
Expertise reversal	Information that is essential for novices may be redundant for experts reversing the relative effectiveness of instructional designs
Problem completion	Similar to the worked example effect based on partial worked examples and can be used during guidance fading
Guidance fading	Due to expertise reversal, as expertise increases, the guidance provided by worked examples should be decreased and eventually eliminated
Imagination	With sufficient expertise, imagining procedures or concepts can be more effective than studying
Transient information	The use of technology can transform permanent into transient information increasing extraneous cognitive load

3.2.2. The Isolated Elements Effect

While the variability effect is due to instructional procedures increasing intrinsic cognitive load, the isolated elements effect is due to instructional procedures decreasing intrinsic cognitive load. Assume that what students are required to learn is very high in element interactivity due to intrinsic cognitive load. It may be so high that the number of elements that must be processed exceeds working memory capacity. In this case, understanding and learning cannot proceed until levels of expertise are attained that permit interacting elements to be incorporated into schemas and treated as single elements using the environmental organizing and linking principle. It may be preferable to initially present the interacting elements in isolated form so that they can be processed even though they cannot be fully understood. Each element can be presented without reference to the other interacting elements. Once learned, the material can be presented again, but on this occasion in fully interacting rather than isolated form so that students can learn the interactions. Pollock, Chandler, and Sweller (2002) presented students with very complex information in isolated elements form thus reducing the intrinsic cognitive load followed by a presentation of the same information with the links between elements indicated. Another group was presented with the fully interacting material twice. The students who were presented with the elements in isolated form first performed better on subsequent test problems, providing an example of the isolated elements effect.

3.3. Extraneous cognitive load

Just as element interactivity determines intrinsic cognitive load, it also determines extraneous cognitive load. While the interacting elements that generate an intrinsic cognitive load are unavoidable other than by changing the task or levels of expertise, extraneous load is under the control of instructors and so the interacting elements due to extraneous cognitive load can be reduced or eliminated by changing instructional procedures. Some instructional procedures require learners to unnecessarily process many elements of information simultaneously resulting in a heavy, extraneous cognitive load that interferes with learning. These interacting elements should be eliminated because unlike intrinsic cognitive load, extraneous cognitive load should always be reduced with no conditions under which it should be increased. There are many cognitive load effects based on instructional techniques designed to reduce extraneous cognitive load.

3.3.1. The Goal-Free Effect

This cognitive load effect was the first to be demonstrated and the first to indicate the negative consequences of a means-ends problem-solving strategy (Sweller, 1988; Sweller, Mawer, & Ward, 1983). The effect occurs when students who are provided problems without a conventional goal outperform students presented with conventional problems on subsequent tests. A goal-free problem will require students to, for example, "calculate the value of as many variables as you can" or "calculate the value of as many angles as you can" rather than, for example, "How fast was the car traveling?" or "What is the value of angle ABC?"

In order to solve a conventional problem, learners must use a means-ends problem-solving strategy in which they consider both the current problem state and the goal state, find differences between the current problem state and the goal state, and find problem-solving operators to reduce these differences. The many interacting elements associated with this process impose an extraneous cognitive load that can overwhelm working memory and interfere with learning. In contrast, goal-free problem solving only requires learners to consider their current problem state and any operator that can alter that state. The reduction in extraneous working memory load due to the reduction in the number of interacting elements by the use of goal-free problems increases the information transferred to the long-term memory store.

While the goal-free effect is an interesting effect, goal-free problems can only be used under conditions where calculating as many variables as possible results in the calculation of a limited number of instructionally relevant variables. Some problems meet this requirement but many do not. For this reason, the worked example effect, discussed next, was devised as a universal procedure.

3.3.2. The Worked Example Effect

The worked example effect (Renkl, 2005) is probably the best known among the cognitive load theory effects. It is demonstrated when students learn more by studying a problem and its solution rather than solving the problem themselves. For example, learners may be presented with the problem, $(a + b)/c = d$, solve for a, for which they are required to find a solution. This problem-solving condition can be compared with a worked example condition in which learners are presented with the same problem along with its worked solution:

$$(a + b)/c = d$$
$$a + b = dc$$
$$a = dc - b$$

The worked example effect is demonstrated when the worked example condition performs better on subsequent problem-solving tests.

Since its demonstration by Sweller and Cooper (1985), the worked example effect has been replicated on a large number of occasions. It occurs because, as is the case with the goal-free effect, problem-solving search is associated with a large number of interacting elements that generate a heavy extraneous cognitive load. In contrast, studying a worked example reduces the number of interacting elements that need to be processed in working memory. Consider a person who does not have a solution schema and so cannot use the environmental organizing and linking principle when attempting to solve the above problem, $(a + b)/c = d$, solve for a. To solve the problem, a sequence of moves that will isolate a must be found. The addend b and the denominator c must be removed from the left side of the equation. What procedure could be used to remove b? Using the randomness as genesis principle, either subtracting b or dividing by b might work. In fact, neither of these procedures seems possible. Perhaps attending to c might work. As can be seen, there are a large number of elements that must be considered when searching for a problem solution. When knowledge is unavailable, the randomness as genesis principle must be used. In contrast, the elements required to study the worked example are all essential for someone who is learning to solve this category of problems and do not extend beyond the elements incorporated in the example. The randomness as genesis principle is not required to generate moves, the use of worked examples reduces the number of interacting elements associated with solving a problem and so reduces extraneous cognitive load, facilitating learning as indicated on subsequent test problems.

Most of the early studies on the worked example effect used curriculum materials from mathematics, science, and other technical areas. The problems were well defined as is common with problems in these areas. There are no cognitive reasons why the worked example effect should not be equally effective in areas that usually deal with ill-defined problems. There are now an increasing number of demonstrations of the worked example effect in ill-defined areas associated with language-based curricula or design issues. For example, Rourke and Sweller (2009) demonstrated the worked example effect when teaching learners to recognize furniture designers' styles.

The worked example effect follows closely from the principles used above to describe human cognitive architecture. Studying worked examples allows us to accumulate the large number of schemas associated with skill in an area in accordance with the information store principle. These schemas are best acquired by borrowing information provided by others in accordance with the borrowing and reorganizing principle. We learn more slowly if we attempt to acquire the same information by problem solving via the randomness as genesis principle. The narrow limits of change principle ensures that reducing working memory load by presenting learners with worked examples rather than having them solve problems facilitates learning. Once problem-solving schemas have been stored in long-term memory, we can solve problems that we otherwise would have great difficulty in solving as indicated by the environmental organizing and linking principle.

3.3.3. The Split-Attention Effect

A large number of other cognitive load theory effects are related to the worked example effect. The split-attention effect was the first of those satellite effects. Worked examples can be effective provided they reduce the need to process interacting elements that are extraneous to learning. Algebra worked examples of the type exemplified above do reduce extraneous cognitive load. Nevertheless, if a worked example is structured in a manner that does not reduce extraneous cognitive load, it will not be effective. Consider a typical geometry worked example. It usually consists of a diagram and a set of statements next to or under the diagram that indicate geometric relations such as "Angle ABC = Angle XBZ (vertically opposite angles are equal)." In order to understand this statement, learners must search for the two relevant angles using the randomness as genesis principle since the angles could be anywhere on the diagram. Search, as indicated above, can be expected to involve a large number of elements of information and processing these elements imposes an unnecessary working memory load—an extraneous cognitive load. That extraneous cognitive load is imposed because learners must split their attention between the diagram and the statements. Alternatively, if the statements are placed at appropriate locations on the diagram or if arrows link the statements with appropriate diagram locations, a search for referent locations no longer is necessary, reducing extraneous cognitive load due to the elimination of the need to use the randomness as genesis principle to process the statements. We might expect such physically integrated worked examples to be superior to conventional, split-attention versions.

Comparing worked examples presented in a split-attention format with a physically integrated format indicates that the integrated format facilitates learning (Sweller, Chandler, Tierney, & Cooper, 1990; Tarmizi & Sweller, 1988; Ward & Sweller, 1990). The effect is relevant to all forms of instruction, not just worked examples. Any instructional procedure including initial instruction prior to the presentation of worked examples and including forms of instruction other than diagrams and text such as multiple sources of text, multiple diagrams, or even physical equipment such as computers (Sweller & Chandler, 1994) should be analyzed from the perspective of the split-attention effect with the aim of physically integrating split-attention materials so that learners do not have to mentally integrate them. Ayres and Sweller (2005) provide a review of the split-attention effect.

It needs to be noted that the split-attention effect applies only to sources of information that are unintelligible in isolation. In order to understand a diagram and a text, for example, both should only become intelligible once they have been physically or mentally integrated. If they do not have to be integrated in order to be understood because, for example, the text merely redescribes the diagram, there are no grounds for assuming that the split-attention effect applies. Under these circumstances, the diagram or text may be redundant, leading to the redundancy effect described below.

3.3.4. The Modality Effect

This effect is closely related to the split-attention effect. When faced with two sources of information that cannot be understood in isolation, rather than physically integrating the two sources, they can be presented in different modalities. One source can be presented visually, while the other source can be presented aurally. Dual modality presentation should increase effective working memory and so decrease cognitive load.

There are theoretical grounds for suggesting that dual modality presentation should increase effective working memory capacity. According to Baddeley's model (Baddeley, 1999), working memory includes an auditory loop for processing speech and a visual–spatial sketchpad for processing visual material. These two processors are partially independent and both are limited in capacity. By using both, working memory capacity should increase (Penney, 1989).

Consider again, geometry instruction presented entirely visually with text presented in written rather than spoken form. The visual channel must be used to process diagrams and must also be used to initially process the written text. The written text then will need to be converted into auditory form for further processing. The visual channel has a limited processing capacity and so it can readily be overloaded. The need to initially process the written text using the visual channel and then to convert the written text into auditory text can be expected to impose an extraneous cognitive load that can interfere with the transfer of information to long-term memory. As an alternative, assume that the written text is presented in spoken rather than written form. The visual channel is no longer needed to process the text nor is there a need to convert the information into auditory form for further processing. The auditory channel only needs to be used to process spoken text. The consequence should be a reduction in the cognitive load imposed on the visual channel that can be expected to enhance learning.

This hypothesis was first tested by Mousavi, Low, and Sweller (1995) using geometry problems. They obtained the modality effect with students performing better on subsequent tests after learning using an audio-visual format rather than a visual only format. These results have been replicated on many occasions (see Ginns, 2005, for a meta-analysis). While the effect is very robust, there are many conditions under which it is known not to occur, with many of these conditions leading to new cognitive load effects. The modality effect will occur only under the same conditions required for the split-attention effect. The two sources of information must be unintelligible in isolation. If text, for example, merely restates the information in a diagram, it will lead to redundancy, not to the modality effect. In addition, the effect will not be obtained if intrinsic cognitive load is low due to the element interactivity effect. Neither will the effect be obtained if levels of expertise are high, due to the expertise reversal effect. Finally, in a very recent work, it was indicated that if text is lengthy and complex, a reverse modality effect is obtained due to the transient information effect. Each of these effects is separately discussed below.

3.3.5. The Redundancy Effect

The redundancy effect occurs when the addition of redundant information interferes with learning. The effect can be obtained using sources of information that on the surface appear similar to those that lead to the split-attention effect. The distinction between the two effects derives from the relation between the multiple sources of information. In the case of the split-attention effect, the multiple sources of information are unintelligible in isolation and must be integrated, mentally or physically, before they can be understood. In the case of redundancy, the sources of information are intelligible in isolation and do not need to be integrated in order to be understood. For example, a text may merely redescribe a diagram that is intelligible in its own right. Such text is redundant. The redundancy effect occurs when any additional information is presented that is not required. Often, but not always, the redundant information redescribes other information. Redundant information is defined as any unnecessary information.

The redundancy effect is caused by the introduction of unnecessary interacting elements resulting in an extraneous cognitive load. For example, if learners are presented with a self-explanatory diagram along with text that redescribes the diagram, they will attempt to process both the elements that constitute the diagram and the elements that constitute the text. They are likely to attempt to relate the diagram and the text. Such attempt to relate diagrams and text is likely to unnecessarily require the use of random generate and test via the randomness as genesis principle. The additional elements that need to be processed in working memory introduce an extraneous cognitive load.

The redundancy effect was first demonstrated using diagrams and redundant text (Chandler & Sweller, 1991). A diagram alone demonstrating the flow of blood in the heart, lungs, and body resulted in more learning than the same diagram with text redescribing the diagram. The effect has been replicated many times using a variety of materials other than diagrams and text. For example, learning to use machinery such as computers can be facilitated by the use of diagrams without the presence of the computer (Sweller & Chandler, 1994). As another example, verbal material should not be presented simultaneously in spoken and written form (Kalyuga, Chandler, & Sweller, 2004). There are many other examples. A review of the redundancy effect may be found in Sweller (2005).

3.3.6. The Element Interactivity Effect

All cognitive load effects rely on the information that is being processed imposing a heavy, intrinsic cognitive load. The information must be complex. If element interactivity due to intrinsic cognitive load is low, any element interactivity due to extraneous cognitive load may have few instructional consequences. It may be possible to process the interacting elements due to extraneous cognitive load without exceeding working memory capacity. If so, cognitive load effects will not be obtained when element interactivity due to intrinsic cognitive load is low. Information can be processed in working memory and transferred to the long-term store even under the presence of elements imposing an extraneous cognitive load. Neither the split-attention nor the redundancy effects are likely to be obtained using intrinsically low element interactivity information (Sweller & Chandler, 1994). Similarly, the modality effect is unlikely to be obtained with such material (Tindall-Ford, Chandler, & Sweller, 1997) along with several other cognitive load effects (Leahy & Sweller, 2005, 2008).

3.3.7. The Expertise Reversal, Problem Completion, and Guidance Fading Effects

The element interactivity effect is concerned with changes in the complexity of information presented to learners. The expertise reversal effect, in turn, is concerned with changes in learners' levels of expertise. The two effects are complementary because the complexity of information and the levels of expertise can compensate for each other with opposing effects on element interactivity. According to the environmental organizing and linking principle, increases in knowledge result in decreases in element interactivity and complexity as interacting elements are incorporated into schemas that are treated as a single element. Thus, intrinsic element interactivity can be decreased either by changing to a task with lower element interactivity or by increasing levels of learner expertise.

The expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003) occurs under the following conditions. Assume that instructional procedure A is superior to instructional procedure B using novice learners. With increasing expertise, the difference between the two procedures narrows and then disappears, before reappearing as a reverse effect with instructional procedure B proving superior to instructional procedure A. The expertise reversal effect has been demonstrated with many cognitive load effects including the worked example (Kalyuga, Chandler, Tuovinen, & Sweller, 2001), split-attention (Kalyuga, Chandler, & Sweller, 1998), and modality effects (Kalyuga, Chandler, & Sweller, 2000).

The expertise reversal effect relies on the redundancy effect. The inclusion of material that may be essential for novices to understand information may be redundant for more knowledgeable learners and so interfere with rather than facilitate learning. Consider the worked example effect. For novices, studying worked examples may facilitate learning compared to solving the equivalent problems. Searching for problem solutions increases extraneous element interactivity that interferes with learning. As expertise increases, learners may still need additional practice but may have sufficient knowledge to no longer need to search for solutions. It may be easier for them to generate a problem solution rather than study a solution provided by someone else. For example, most readers of this chapter are likely to find it easier to solve the problem $(a + b)/c = d$, solve for a, rather than study a worked example. Studying a worked example is likely to be redundant and so increase rather than decrease working memory load. As a consequence, for novices who have just begun to learn algebra, it may be easier to study a worked example than solve the equivalent problem; while for more knowledgeable learners, it may be easier to solve the problem than study the equivalent worked example, resulting in an expertise reversal effect. This effect was demonstrated by Kalyuga et al. (2001) using worked examples.

It follows from the worked example effect that novices should initially be presented with worked examples to study. With increasing expertise, these worked examples should be replaced by problems. Initially, worked examples can be replaced by completion problems that include part of the solution with the rest to be completed by learners (Paas, 1992; van Merrienboer, 1990). The *completion effect* is similar to the worked example effect and occurs when students presented with completion problems learn more than students presented with full problems. With further increases in expertise, completion problem may be replaced by full problems. This process of fading worked examples is superior to either just solving problems or just studying worked examples and is known as the *problem fading effect* (Salden, Aleven, Schwonke, & Renkl, 2010).

3.3.8. The Imagination Effect

This effect is also subject to expertise reversal. It occurs when learners are asked to imagine a concept or procedure rather than study it (Cooper, Tindall-Ford, Chandler, & Sweller, 2001). For example, learners may be presented with a worked example of an algebra problem. Rather than being asked to study the worked example, learners under *imagination* conditions are asked to look at the example and then turn away from it and try to imagine the solution to the problem. The imagination effect occurs when imagining a concept or procedure is superior to studying the relevant material. In order to imagine a concept or procedure, it is necessary to process the information in working memory. Novices may have difficulty in processing all the required interacting elements in working memory and so imagining concepts or procedures may be difficult or even impossible. More knowledgeable learners may be able to imagine information more readily because many of the interacting elements are already incorporated into schemas via the environmental organizing and linking principle. As a consequence, the effect can be demonstrated only if levels of expertise are sufficiently high. For novices, studying the information tends to be superior to imagining it because imagining all the necessary interacting elements may overload working memory.

3.3.9. The Transient Information Effect

This effect is a new cognitive load effect. The use of educational technology sometimes has unintended cognitive load consequences. For example, a frequent side effect of using technology is that previously permanent information that can be repeatedly and easily accessed becomes transient and can only be reaccessed with difficulty or cannot be accessed at all. Information is transient if it disappears with the passage of time. Shifting from permanent written text to transitory auditory text or from permanent sets of diagrams to animation provides examples. Auditory information or most animated information disappears as new information is presented and so is transitory. If the information being conveyed is high in element interactivity, presenting it in transient form can have negative consequences. Having to remember previous, high element interactivity information that is no longer available and integrate it with currently appearing information can severely overload working memory.

Evidence for this hypothesis was obtained by Leahy and Sweller (in press) when testing for the modality effect. They ran two experiments comparing dual modality with visual only presentations. Primary school students were taught how to interpret time/temperature graphs showing the variations in temperature during the day. The first experiment included relatively lengthy, complex spoken statements such as "Find 35C on the temperature axis and follow across to a dot" while referring to a graph. The second experiment provided exactly the same information except that the statements were divided into smaller segments. The above statement, for example, was divided into "Find 35C on the temperature axis" and "Follow across to a dot." The first experiment with the longer statements demonstrated a reverse modality effect with the visual only material that included written statements proving superior to the audiovisual presentation. The second experiment with the shorter statements indicated a conventional modality effect.

The Leahy and Sweller work was not the first to obtain a reverse modality effect. Tabbers, Martens, and van Merriënboer (2004) also obtained a reverse modality effect using relatively lengthy, complex verbal information. These results can be explained readily from a cognitive load theory perspective. Assume that learners are faced with a relatively complex statement such as "Find 35C on the temperature axis and follow across to a dot." Holding this statement in working memory while referring to a graph may overload working memory. If presented in spoken form, the entire statement will need to be held and processed in working memory. In contrast, if it is presented in written form, learners can easily divide and return to the statement in part or in whole whenever they need to. For example, they can quickly scan the entire statement once and then return to the first clause, "Find 35C on the temperature axis...," process that statement with respect to the graph by finding the 35C point, and then return to the statement to process the rest of the statement "...and follow across to a dot." If presented in auditory form, learners would need to have memorized the entire statement using the information store and environmental organizing and linking principles in order to engage in a similar activity. Accordingly, a visual text along with a visual diagram is superior to an audiovisual presentation.

If the statements are presented in shorter form, they are likely to be automatically held in working memory irrespective of whether they are presented in spoken or in written form. For shorter statements, the expansion of working memory due to the use of both auditory and visual channels should result in the conventional modality effect obtained in a large number of studies over many years (Ginns, 2005).

The transient information effect should apply equally to any transient information such as complex, high element interactivity animations. Preliminary results confirm that the length of animations can determine their relative effectiveness compared to static graphics.

3.4. Summary of element interactivity and the cognitive load effects

Element interactivity is central to cognitive load theory and the cognitive load effects. When we must process multiple, interacting elements in working memory simultaneously, an excessive or inappropriate cognitive load may be generated. If cognitive load is intrinsic to the information being assimilated as it occurs for the variability and isolated elements effects, it needs to be altered. Altering intrinsic cognitive load will alter what is learned and understood. Intrinsic cognitive load cannot be altered if what needs to be learned is unaltered and if levels of expertise remain the same.

The vast majority of cognitive load effects are due to a reduction of extraneous cognitive load. If instructional procedures require learners to unnecessarily process interacting elements because of the manner in which information is presented, especially if the presentation of information requires learners to use the randomness as genesis rather than the borrowing and reorganizing principle, extraneous cognitive load will be high and should be reduced. A reduction in extraneous cognitive load will permit working memory resources to be mobilized to deal with intrinsic load that is germane to learning. Extraneous cognitive load can be reduced by altering instructional procedures as indicated in this section (see Table 2).

4. Conclusions

The cognitive load effects provide the ultimate justification for cognitive load theory. Nevertheless, they should not be considered in isolation. Human cognitive architecture and the categories of cognitive load are essential. There has been a tendency for some in the field to assume that the cognitive load effects can be considered in isolation from the cognitive architecture that gave rise to the effects. This view is misguided. We cannot automatically assume that, for example, studying worked examples is superior to solving problems or presenting information in a split-source format is worse than presenting information in an integrated format. None of the effects should be considered in isolation from the theoretical constructs that gave rise to them. Studying worked examples is frequently superior to solving problems but only if extraneous cognitive load is reduced. If it is not reduced because, for example, worked examples are presented in split-source format or student knowledge is sufficiently high to not require worked examples, then the use of worked examples will be ineffective. Similarly, while we know that dual mode presentations of information can be very effective, we also know that if verbal information is redundant, using a dual mode presentation will not be effective because the redundant information increases extraneous cognitive load. We now also know that lengthy, complex, high element interactivity verbal material needs to be presented in written, not spoken form. The modality effect does not provide an excuse to use audiovisual presentations irrespective of other cognitive load factors. If dual modality presentation leads to a heavy, extraneous cognitive load as will happen if lengthy, complex statements are presented in auditory form, we should not expect to obtain a modality effect. A reverse modality effect is more likely.

The cognitive architecture of Section 2 can be used to assess the likely effects of any instructional intervention. According to this architecture, the purpose of instruction is to increase usable knowledge held in long-term memory via the information store principle. This knowledge allows our working memory to function at a high level according to the environmental organizing and linking principle, permitting us to engage in activities that otherwise would be difficult or impossible. Obtaining information from others is the best way of acquiring knowledge according to the borrowing and organizing principle. If knowledge is not held in long-term memory, we must process information in working memory that is limited in capacity and duration when dealing with novel information according to the narrow limits of change principle. We can acquire novel information while problem solving in accord with the randomness as genesis principle, but that process requires working memory resources that consequently are unavailable for learning. Instructional procedures that do not meet the objective of increasing knowledge in long-term memory while decreasing any unnecessary load on working memory are likely to be ineffective.

We should never ignore human cognitive architecture when designing instruction. It is not an optional extra.

REFERENCES

Ayres, P., & Sweller, J. (2005). The split-attention principle. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 135–146). New York: Cambridge University Press.

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

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewoods Cliffs, NJ: Prentice Hall.

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

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–75). Hillsdale, NJ: Erlbaum.

Chiesi, H., Spilich, G., & Voss, J. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behaviour*, *18*, 257–273.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology: Applied*, *7*, 68–82.

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

De Groot, A. (1965). *Thought and choice in chess.* The Hague, Netherlands: Mouton (original work published in 1946).

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

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*, 725–747.

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.

Geary, D. (2008). An evolutionarily informed education science. *Educational Psychologist*, *43*, 179–195.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*, 313–331.

Jablonka, E., & Lamb, M. J. (1995). *Epigenetic inheritance and evolution.* New York: Oxford University Press.

Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life.* Cambridge, MA: MIT Press.

Jeffries, R., Turner, A., Polson, P., & Atwood, M. (1981). Processes involved in designing software. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 255–283). Hillsdale, NJ: Erlbaum.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23–31.

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

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 126–136.

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors*, *46*, 567–581.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, *93*, 579–588.

Kirschner, P., Sweller, J., & 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. *Educational Psychologist*, *41*, 75–86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*, 661–667.

Kotovsky, K., Hayes, J., & Simon, H. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*, 248–294.

Leahy, W., & Sweller, J. (2005). Interactions among the imagination, expertise reversal, and element interactivity effects. *Journal of Experimental Psychology: Applied*, *11*, 266–276.

Leahy, W., & Sweller, J. (2008). The imagination effect increases with an increased intrinsic cognitive load. *Applied Cognitive Psychology*, *22*, 273–283.

Leahy, W., & Sweller, J. (in press). Cognitive load theory, modality of presentation and the transient information effect. Applied Cognitive Psychology.

Luchins, A. (1942). Mechanisation in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*, (Whole No. 248).

Luchins, A. (1942). Mechanisation in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*, (Whole No. 248).

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

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, *87*, 319–334.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434.

Paas, F., & van Merrienboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122–133.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*, 398–422.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198.

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

Portin, P. (2002). Historical development of the concept of the gene. *Journal of Medicine and Philosophy*, *27*, 257–286.

Renkl, A. (2005). The worked out example principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229–245). New York: Cambridge University Press.
Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction*, *19*, 185–199.
Salden, R., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, *38*, 289–307.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.
Simon, H., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology*, *5*, 29–46.
Stotz, K., & Griffiths, P. (2004). Genes: Philosophical analyses put to the test. *History and Philosophy of the Life Sciences*, *26*, 5–28.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285.
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, CA: Academic Press.
Sweller, J. (2005). The redundancy principle. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 159–167). New York: Cambridge University Press.
Sweller, J. (2009). Cognitive bases of human creativity. *Educational Psychology Review*, *21*, 11–19.
Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, *22*, 123–138.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory.* New York: Springer.
Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*, 185–233.
Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, *119*, 176–192.
Sweller, J., & Cooper, G. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*, 59–89.
Sweller, J., & Gee, W. (1978). Einstellung, the sequence effect, and hypothesis theory. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 513–526.
Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, *112*, 639–661.
Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, *4*, 434–458.
Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, *74*, 71–81.
Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, *80*, 424–436.
Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, *3*, 257–287.
van Merrienboer, J. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research*, *6*, 265–285.
Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 7, 1–39.
West-Eberhard, M. (2003). *Developmental plasticity and evolution.* New York: Oxford University Press.