Chapter 12
Cognitive Load Theory, Evolutionary Educational Psychology, and Instructional Design

John Sweller

Cognitive load theory (Sweller, Ayres, & Kalyuga, 2011) is an instructional approach based on our knowledge of human cognitive architecture, including the limits of working memory, the organization of information in long-term memory, and the interactions between these memory systems. That architecture is used to generate novel instructional procedures intended to facilitate learning in educational settings. Once an instructional procedure is developed based on this theory, its effectiveness is tested by comparing learning outcomes to more traditional procedures using randomized controlled trials. When those learning outcomes favor the new instructional procedure, a new cognitive load effect has been identified for further study and a potential new instructional procedure is available for use in the classroom.

Those aspects of human cognitive architecture relevant to instruction and used by cognitive load theory depend on evolutionary educational psychology in two respects. First, biological evolution can be used to determine categories of knowledge that are important to instructional considerations. Second, the selection pressures that drive evolution by natural selection are analogous to those that operate during human learning. I will begin by considering human cognitive architecture from an evolutionary educational psychology perspective, and then link these to instructional design.

Evolutionary Educational Psychology and Human Cognition

Early versions of cognitive load theory did not use evolutionary educational psychology when discussing human cognitive architecture, but instead placed the primary emphasis on relations between working and long-term memory.

J. Sweller (✉)
School of Education, University of New South Wales, Sydney, NSW 2052, Australia
e-mail: j.sweller@unsw.edu.au

While those relations still are critical to the theory, the subsequent emergence of a viable evolutionary educational psychology placed the relations between working and long-term memory into a context that provided substantially more explanatory power and generated a wider range of hypotheses. By using evolutionary educational psychology, the categories of knowledge to which cognitive load theory did and did not apply became clearer, as did the way in which information was processed, stored, and used during and subsequent to instruction.

Categories of Knowledge

Knowledge probably can be categorized in an infinite number of ways, but for present purposes, the only categories that matter are ones that have instructional implications. Categorization schemes in which the same instructional procedures are equally effective across the identified categories have minimal or no instructional implications. For example, if the same instructional techniques are important in teaching concepts and teaching procedures, the distinction between concepts and procedures becomes irrelevant from an instructional perspective, even if it is important from other perspectives. One scheme based on evolutionary educational psychology was devised by David Geary and has profound significance for instructional procedures (Geary, 2012, this volume). He divided knowledge into biologically primary and secondary knowledge, two categories that require vastly different experiences for their development and so different instructional procedures.

Biologically primary knowledge. We have evolved to acquire biologically primary knowledge over countless generations. It tends to be knowledge that is critical to our survival and is organized around the domains of folk psychology such as social abilities, folk biology such as knowledge of other species, and folk physics such as the ability to navigate from place to place (Geary, 2005). Recognizing faces, learning to listen to and speak a first language, basic social skills associated with relationships are all features of folk psychology, for instance. We also have evolved general problem-solving skills and the ability to plan ahead and strategize.

Biologically, primary knowledge has several important characteristics. First, it tends to be modular with, for example, the ability to recognize faces likely to have evolved independently and in a different epoch than language skills. Thus, the manner in which we acquire one skill may differ markedly from the manner in which we acquire a different, unrelated skill. Second, because we have evolved to acquire biologically primary knowledge, it tends to be acquired easily, automatically, and unconsciously through natural activities, such as play and social discourse. To acquire biologically primary skills, we merely need membership of a functioning (or in some cases, even a dysfunctional) society. As a consequence, biologically primary skills do not need to be explicitly taught, or indeed, taught at all. All normally functioning individuals will acquire those skills.

Third, it is likely that most of the generic skills considered important in education are biologically primary (Tricot & Sweller, 2014). For example, because of their

importance in many real-world contexts, we may have evolved general problem solving and planning skills. Means-ends analysis provides an example of a general problem-solving skill (Newell & Simon, 1972). We solve many novel problems by noting our current and goal problem states, finding problem-solving operators to reduce differences between the two states, and then repeating the process until the goal has been attained. This means-ends strategy is a generic skill that is commonly used, but without any evidence that it is teachable. It constitutes a complex, primary skill that all normal humans acquire without instruction. It follows that such generic skills do not need, indeed cannot, be taught because they are automatically acquired. Including instruction of such skills in curricula is likely to be futile.

Biologically secondary knowledge. In contrast to biologically primary knowledge that we all must acquire in order to function appropriately in any society, biologically secondary knowledge is culturally specific. While the knowledge itself is entirely domain-specific, we have evolved to acquire any secondary knowledge generically. In other words, the ability to acquire secondary knowledge is biologically primary (Geary, 2005). We do not need to be taught how to obtain biologically secondary knowledge because we have evolved to do so. As a result, teaching learners how to develop knowledge as opposed to teaching them the actual knowledge may be a pointless exercise. The manner in which we acquire biologically secondary knowledge is largely identical irrespective of the nature of that knowledge: We have evolved to acquire a wide variety of types of biologically secondary knowledge in a similar manner.

Examples of biologically secondary knowledge can be found in every curriculum area found in any educational establishment. We invented schools in order to teach biologically secondary knowledge because, unlike primary knowledge, it is unlikely to be acquired without the functions and procedures found in educational establishments.

There are two characteristics of biologically secondary knowledge that are critical to instructional issues. First, it is domain-specific (Tricot & Sweller, 2014). To learn to solve mathematics problems, we do not need to be taught generic, cognitive problem-solving skills, such as means–ends analysis. These skills are already part of our evolved repertoire, although some domain-specific problem-solving procedures such as the use of formal logic and experimental design must be explicitly taught in the areas to which they are applied (but see Gray, this volume; Toub et al., this volume). For example, the experimental designs suitable for biology or psychology bear little resemblance to those used in physics. Learning those procedures is a biologically secondary task that must be taught explicitly. Similarly, we do need to be taught the procedures required to solve particular, narrow classes of problems. For example, we need to learn how to solve problems of the type, $a/b = c$, solve for a. The knowledge gained is domain-specific in that knowing how to solve this category of problem will not be of assistance in solving unrelated mathematics problems or unrelated, non-mathematical problems.

The second important characteristic of biologically secondary knowledge is that, unlike biologically primary knowledge, it can be difficult to learn, requires conscious effort, and is learned much more easily with explicit instruction rather than

minimal guidance (Kirschner, Sweller, & Clark, 2006). When acquiring biologically primary knowledge, learners can be left to their own devices because they have evolved to acquire such knowledge. It is inadvisable to provide minimal guidance when dealing with biologically secondary knowledge. Without guidance, the information may be misunderstood or not acquired at all, a risk that is minimal when dealing with biologically primary knowledge.

Natural Information Processing systems

From the above analysis, the major function of instruction is to assist learners to acquire biologically secondary knowledge. The cognitive architecture associated with the acquisition and use of biologically secondary knowledge is closely analogous to the process of biological evolution itself. The suggestion that evolution by natural selection and human cognition is analogous has a very long and illustrious history (Campbell, 1960; Darwin, 1871/2003; Popper, 1979; Siegler, 1996). Both human cognitive architecture and evolution by natural selection are examples of natural information processing systems (Sweller & Sweller, 2006). They can be described using five basic principles.

Information store principle. Natural information processing systems require a very large store of information in order to function in a natural environment. In the case of biological evolution, that store is represented by a genome. While there is no agreed upon measure of the size of a genome, any measure considered results in thousands of units of information for the smallest genomes and much more for larger genomes (Portin, 2002; Stotz & Griffiths, 2004).

For human cognitive architecture, long-term memory provides the functional equivalent of a genome. Competent performance in any substantive, biologically secondary area requires many years of deliberate practice to improve performance (Ericsson, Krampe, & Tesch-Romer, 1993). That practice results in the storage of large amounts of domain-specific information. The initial evidence for the huge amounts of information stored in long-term memory came from De Groot's (1965) classic work on chess. He found that chess masters did not engage in more problem-solving search than weekend players. The only difference between the two groups was in memory of chessboard configurations. Chess masters, who have shown a configuration taken from a real game for 5 s, were able to accurately replace over 80 % of the pieces. Weekend players only were able to replace less than 30 % of the pieces. Chase and Simon (1973) replicated these results and in addition found no difference between masters and weekend players' presented random board configurations as opposed to real game configurations. For random configurations, accuracy was similar to that of weekend players' presented configurations taken from real games. Thus, only chess masters presented real game configurations performed at a high level. Similar results have been obtained in a variety of areas relevant to education, including learning algebra and computer programming (e.g., Egan & Schwartz, 1979; Jeffries, Turner, Polson, & Atwood, 1981; Sweller & Cooper, 1985).

The work on expertise and particularly De Groot's (1965) work changed our view of human cognition and, indeed, our view of ourselves. Arguably, it is the most important finding of cognitive psychology. Until this work, we saw the defining characteristic of human cognition to be our ability to "think," but a definitive definition has remained elusive. The new role of long-term memory in human cognition, while not providing a definition, set us on the road. Playing chess at master or grand master level surely required thought and it turned out that long-term memory was critical to that thought to an extent that previously had not been imagined.

With respect to learnable factors as opposed to inherited factors, a key difference between someone who is good at an intellectual activity in a specific secondary domain and someone who is not seems to be largely dependent on the information held in long-term memory. In this context, we know, for example, that working memory capacity is dramatically affected by the contents of long-term memory (see the organizing and linking principle below) and that IQ tests need to be re-standardized every few years and show a continuously rising trend (Flynn, 1987). We also know that one additional year of schooling increases IQ by more than one additional year of age (Cahan & Cohen, 1989). A parsimonious explanation of changes in working memory capacity and IQ can be provided by assuming that both are strongly affected by the contents of long-term memory. Indeed, at present, there is no clear evidence of any other factor being relevant.

Whether dealing with a genome or long-term memory, the information held in the information store is central to natural information processing systems. Natural environments tend to be complex. To deal with that complexity, a large store of information is essential.

Borrowing and reorganizing principle. How is the large amount of information held in a natural information store acquired? The manner in which an individual genome obtains its information is well-known. During either sexual or asexual reproduction, information is borrowed from ancestors. In the case of sexual reproduction, that information is necessarily reorganized as an essential part of the process.

An analogous process is used by human cognition. We imitate what others do (Bandura, 1986), we listen to what they say, and we read what they write. We are one of a very small number of species that have evolved to provide and receive information via deliberate teaching from other members of the species (Thornton & Raihani, 2008). Our ability to obtain information from other people is biologically primary, even when used to acquire a biologically secondary skill such as reading. The skill is secondary, but the general ability to obtain the information required for that secondary skill is primary. We have to teach people to read, but we do not have to teach them to obtain information by reading because once one is taught how to read, the skill can tap into our biologically primary natural language and social-information systems. People know that they can obtain information from other people by reading because that knowledge is biologically primary and does not need to be taught.

The information we obtain from others is reorganized in the same manner as information is reorganized during sexual reproduction. Knowledge obtained from other people is automatically combined with knowledge already held in long-term

memory to provide new knowledge that may be unique and useful. For this reason, the information obtained from other people is rarely recorded precisely. It is constructed when combined with previously held knowledge.

From an instructional perspective, it follows that instruction should provide learners with information. Cognitive load theory places its major emphasis on techniques designed to facilitate the acquisition of domain-specific, biologically secondary knowledge using explicit instruction.

Randomness as genesis principle. While we have evolved to obtain most of our knowledge from other people, that knowledge needs to have been created in the first place. Evolution by natural selection also needs to create novel information. It does so by random mutation that is the initial source of all biological variation.

In the case of human cognition, random generate and test during problem solving creates novel concepts and procedures (Sweller, 2009). When presented with a problem, we will attempt to solve it automatically using information held in long-term memory. The bias to use known solution procedures is biologically primary and so unteachable. A known solution always will be used if it is available. If a problem is novel with no known solution stored in long-term memory, it may be possible to generalize from a known solution to a similar problem. Again, if we have access to a problem from which we can generalize, we will do so automatically. Generalizing also is unteachable because it is a biologically primary skill. Of course, if the problem is novel, by definition we cannot know whether a solution to a known problem really does generalize to the new one. We only can find out whether an old solution works on a new problem by trying it out. In a form of generate and test, we generate the solution and see if it works. If it works, we may store the new problem and its solution in long-term memory for use on subsequent occasions.

Frequently, when faced with a novel problem, no solution or even partial solution can be obtained from long-term memory. Either from the start or during problem solving, we may find that there are several possible moves that can be made, but we have no knowledge-based information that will indicate which move we should try. At that point, we will have no choice but to randomly choose a move and test it for effectiveness using a random generate and test procedure. Again, if the move or sequence of moves is effective, we may store it in long-term memory for later use, but jettison it if it proves to be ineffective. In this way, new knowledge is created.

It may be argued that no problem-solving move is ever entirely random and that all such moves have some knowledge attached to them. In a sense, that argument must be correct. If we have no knowledge, we probably not only would be unable to solve the problem, we probably could not even assimilate the meaning of the problem to begin solving it. Nevertheless, the fact that some knowledge always is required does not contradict the randomness as genesis principle. In the same way as random mutation does not occur in a vacuum but only is applied to a current genome, so random generate and test during problem solving always will be applied to a current knowledge base. The fact that there must be organized information already stored prior to the randomness as genesis principle being applied does not eliminate the random component. In the case of problem solving, there inevitably will be some circumstances in which no knowledge is available to discriminate

between alternative problem-solving moves. Under those circumstances, random generate and test is unavoidable. When it occurs, new knowledge is created just as new genetic variations are created by random mutation.

Narrow limits of change principle. The randomness as genesis principle has structural consequences. If new information is to be generated randomly, it needs to be restricted in some way. The need for such a restriction can be seen most clearly in the case of human cognition. Assume that during problem solving, three elements of information need to be combined. If no information is available indicating how they should be combined, then there are $3!=6$ possible permutations of the three elements. Assume instead that there are ten elements that need to be combined. There are $10!=3,628,800$ possible permutations. Using a random generate and test procedure, it will take much longer to determine which permutations are beneficial for ten than three elements. Based on ten elements, a useful permutation that needs to be stored may never be found. For this reason, to ensure that useful, previously stored information is not damaged by a sudden large change, both evolution by natural selection and human cognition require mechanisms that prevent large, rapid changes to the store.

Evolution by natural selection solves this problem by limiting the number of mutations that are likely to occur. The epigenetic system is used to vary the number of mutations that might occur at any given genome location. For example, the level of stress in an environment may alter the number of mutations. Similarly, some sections of a genome may have mutations rates thousands of time higher or lower than other sections. Mutation rates can be very high if diversity is required such as venom used to disable prey (Jablonka & Lamb, 2005). In other words, environmental requirements can result in changes in generation rates of mutations. Nevertheless, large numbers of mutations can jeopardize the integrity of a genome and so mechanisms such as DNA repair are required to constrain mutation rates.

The number of mutations that are retained tends to be low in order to ensure that the organized information stored in a genome is not lost by large, random changes that are likely to be fatal. Genetic change due to random mutation is slow. In effect, very small changes are made and tested for effectiveness. Most of those changes are not adaptive and jettisoned over evolutionary time through differential survival and reproduction. Occasionally, a change is adaptive and retained. The result is a series of very small changes over long periods of time that can slowly improve the adaptation of a genome to an environment without destroying the genome.

In the case of human cognition, working memory plays an analogous role to these genetic changes. New information can be obtained during problem solving, but it is obtained very slowly with the characteristics of working memory constituting the limiting factor. When dealing with novel information, working memory capacity is limited to holding about seven items (Miller, 1956) and processing no more than about four or less items (Cowan, 2001) where processing involves combining, comparing, or relating items in some manner. Not only is the capacity of working memory severely limited when dealing with novel information, the duration that novel information will be retained in working memory is constrained to no more than about 20 s without rehearsal (Peterson & Peterson, 1959). As a consequence

of these limitations of working memory when dealing with new information, changes to the long-term memory store are slow in the same way that changes to a genome are slow.

Environmental organizing and linking principle. While the environment influences changes to the information store, the ultimate purpose of this store is to enable adaptive functioning in a given environment. That purpose is realized through the environmental organizing and linking principle. In the case of biological evolution, the epigenetic system can transform genetic functions. For example, while a person's skin cells and liver cells have identical genotypes, they have vastly different phenotypes. Those differences cannot be caused by genetic factors because, for a given individual, the genetic information in the nucleus of a skin cell is identical to the genetic information in the nucleus of a liver cell. The epigenetic system determines the phenotypic differences by turning genes on or off. Rather than determining where mutations occur and the speed of mutations under the narrow limits of change principle, the epigenetic system can determine the different structures and functions of two types of cells by activating or de-activating particular genes using the environmental organizing and linking principle. It can take large amounts of information from the genome to determine specific structures and functions. Under a different environment, it can use different parts of the available genomic information (different sets of base pairs) to determine different structures and functions.

Similarly, while working memory determines which changes are made to long-term memory, it also determines which information held in long-term memory is used to determine action in a given environment. As is the case for the epigenetic system, working memory can take unlimited amounts of information from the information store, in this case long-term memory, to determine actions appropriate to a given environment. The capacity and duration limits that are necessary when working memory deals with novel information are no longer necessary when it deals with organized information stored in long-term memory (Ericsson & Kintsch, 1995). Working memory has no known capacity or duration limits when dealing with stored information from long-term memory.

Two separate functions of working memory and the epigenetic system. The narrow limits of change and the environmental organizing and linking principles indicate two largely unrelated functions of each of working memory and the epigenetic system. Historically, working memory has been treated as a single system (Atkinson & Shiffrin, 1968), with working memory having the same properties when dealing with novel information from the external environment or familiar information stored in long-term memory. In fact, that unified view of working memory, while attractive in some respects, could not be maintained and for that reason, in the current treatment, working memory has very different properties depending on whether it obtains its information from the environment (the narrow limits of change principle) or from long-term memory (the environmental organizing and linking principle). The distinction is so important that Ericsson and Kintsch (1995) suggested an entirely new structure, long-term working memory to deal with information that is stored in long-term memory and then processed in working memory.

(From a functional perspective, it makes no difference whether we describe two separate structures or a single structure with two separate functions.)

The same issue is relevant to the epigenetic system. It usually is treated as a single system that sometimes affects the number and location of mutations and at other times affects expression or inhibition of information stored in the genome. These two functions are regarded as separate and unrelated in the current treatment, closely analogous to the two functions of working memory. Epigenetically generated changes in the location and rate of mutations are considered under the narrow limits of change principle, while epigenetic factors switching genes on or off are considered under the environmental organizing and linking principle.

Cognitive Load Theory and Instructional Design

This cognitive architecture can be used to devise instructional procedures. In the case of human cognition, the environmental organizing and linking principle allows us to engage in activities that otherwise would be impossible. Those activities depend on us having accumulated large amounts of information in long-term memory via a very limited working memory. Cognitive load theory uses this cognitive architecture to devise instructional procedures. Those procedures generated from the above cognitive architecture have several common characteristics. The two most important are an emphasis on explicit instruction rather than minimal guidance and on the primacy of teaching domain-specific knowledge rather than generic skills. These two recommendations derived from our knowledge of human cognitive architecture will be discussed next.

The Importance of Explicit Instruction

Many instructional theories recommend that students should not be presented direct, explicit information, but rather should be encouraged to find information themselves (Gray, this volume). Inquiry learning, constructivist learning, and problem-based learning provide examples. Ultimately, all derive from discovery learning (Bruner, 1961) and cannot be distinguished from discovery learning or from each other. There is little evidence for the effectiveness of minimal guidance and considerable evidence for the importance of explicit instruction (Kirschner et al., 2006; Klahr & Nigam, 2004; Mayer, 2004; but see Toub et al., this volume).

The cognitive architecture described above explains why explicit instruction is important. Humans obtain the vast bulk of the biologically secondary information held in long-term memory via the borrowing and organizing principle. We have evolved to present and obtain such information from others as a biologically primary skill, as noted. Obtaining information from a teacher or instructor is entirely natural for humans but largely, though not entirely, absent in other animals

(see Berch, this volume). Humans have evolved to learn from others and in ways advocated by proponents of discovery learning. This works well for fleshing out primary knowledge, but not for secondary learning (Geary, 1995, this volume). Given that we have evolved to acquire information from others, recommendations that we should not present explicit information to learners can be seen as little short of bizarre from a cognitive science perspective. These theories arise from people's primary folk psychology, without an understanding that secondary learning is very different from primary learning and what works for the latter does not work well for the former. We have evolved both to teach and to obtain information from teachers.

We also are able to obtain information by discovery learning procedures using the randomness as genesis principle. That machinery is essential when information is required, but there are no other people available to provide that information. While we can and must be able to obtain information in this manner and, indeed, the randomness as genesis principle provides the origin of all biologically secondary information, it is a very slow, difficult, and inefficient process for obtaining information. We are far better at obtaining information using the borrowing and organizing principle. Given a choice between having learners discover information and presenting them with the same information, we should present the information.

The Primacy of Domain-Specific Knowledge

Geary's (1995) distinction between biologically primary and secondary information has implications for the type of information we should be presenting to learners and the skills we should be teaching. Over many years, there has been an increasing emphasis in educational research on teaching generic, cognitive skills (Tricot & Sweller, 2014). These are skills that transcend a particular domain, for example, a general problem-solving skill that improves problem-solving performance irrespective of the domain or metacognitive skills that can improve learning in any area. In one sense, that emphasis is understandable. Generic, cognitive skills are likely to be critical to any cognitive functioning, and indeed, are likely to be far more important than domain-specific skills. Facilitating problem-solving skills that transcend a specific area is likely to be much more important than facilitating problem-solving skill in a narrow, specific domain.

While the importance of generic, cognitive skills explains the emphasis placed on them, there has been a marked lack of success in identifying teachable, learnable, generic cognitive skills. A teachable generic cognitive skill is one that results in improved performance on far transfer tasks that differ from the trained tasks but should, in theory, be improved by the training. An emphasis on far transfer is critical in order to ensure that any performance improvement can be attributed to the acquisition of a generic skill rather than domain-specific knowledge. For example, teaching a generic, cognitive skill and using algebra to provide examples and then testing the extent to which acquisition of the skill improved performance on algebra leaves open the possibility that any improvement may be due entirely to increased knowledge of

algebra rather than increased knowledge of the generic, cognitive skill. If algebra is used to teach the generic skill, any test of the efficacy of learning the skill should use an area unrelated to algebraic skill. Despite many studies over many years, there is minimal evidence available that teaching a generic, cognitive skill improves transfer performance (Ritchie, Bates, & Deary, 2015; Tricot & Sweller, 2014).

We are left with the question as to why there continues to be such a strong emphasis in the field on teaching generic, cognitive skills given that research into teaching those skills failed? In some sense, the answer to this question is straightforward. People could see how easy it was for learners to learn to talk, walk, recognize faces etc., but so difficult to learn subject matter in schools. It followed, they suggested, that the difference in difficulty was due to faulty instructional procedures. If only we used the learning procedures common outside of schools, school learning would be just as easy, natural, and enjoyable as learning outside of school. Explicit teaching is not used to teach people how to listen and talk. If we eliminate explicit teaching of, for example, reading and writing, it will be learned as easily and naturally as listening and talking.

Of course, Geary's (1995) distinction between biologically primary and secondary knowledge explains why some information is acquired easily while other information is difficult to acquire. Because of the importance of generic, cognitive skills, most humans must possess them in order to survive. A skill that is essential to survival is a skill that we are very likely to have evolved to obtain easily and automatically without being taught. Such a skill is a biologically primary skill. If so, the failure to find teachable, learnable, generic, cognitive skills is not because such skills are unimportant, but rather because such skills are so important that most learners will have acquired them without instruction. In contrast, domain-specific skills are largely biologically secondary. They have been created over the past few millennia and do not have the built-in skeletal knowledge that makes primary learning easy and automatic. They are not acquired automatically and should be taught explicitly. We invented schools and other educational institutions precisely because the domain-specific, biologically secondary skills taught were not easily learned without deliberate, explicit instruction.

Some Instructional Effects Generated by Cognitive Load Theory

Cognitive load theory has generated a large number of cognitive load effects. Each effect is based on randomized, controlled experiments comparing a new instructional procedure with more conventional procedures. A cognitive load effect is demonstrated when the new procedure results in superior test performance to the older procedure. All of the hypotheses tested were generated using the above cognitive architecture and assume that effective instruction is explicit and concerned with the acquisition of domain-specific knowledge.

Each cognitive load effect is assumed to be caused by differential levels of element interactivity (Sweller, 2010), a concept that is concerned with the number of

interacting elements that must be processed in working memory. As an example, assume learners are faced with a difficult task such as learning the symbols of the periodic table or some of the nouns of a foreign language. While these tasks are difficult, they do not impose a heavy cognitive load. Each element can be learned independently of every other element and so element interactivity is low resulting in a low working memory load. The task may be difficult, but the intrinsic cognitive load of the task is low. In contrast, other tasks may involve far fewer elements that need to be processed simultaneously in working memory, resulting in high element interactivity and a high intrinsic cognitive load. Balancing a chemical equation provides an example as does solving a problem such as $(a+b)/c=d$, solve for a. To solve this problem, all of the elements must be considered simultaneously because a change in one element is likely to have consequences for every other element. Element interactivity and the intrinsic cognitive load imposed by this task will be high. That intrinsic cognitive load only can be altered by altering the task or by acquiring knowledge stored in long-term memory. With knowledge, the equation, $(a+b)/c=d$, will be treated as a single element rather than multiple elements and so reduce intrinsic cognitive load.

Element interactivity also can be varied by instructional procedures (Sweller, 2010). Some instructional procedures require learners to process many elements simultaneously, while other procedures can substantially reduce the number of elements that need to be processed. Variations in element interactivity due to instructional procedures are referred to as variations in extraneous cognitive load. Most of the effects generated by cognitive load theory depend on a reduction in extraneous load on working memory resources.

The effects only will be very briefly summarized here. More detailed summaries may be found in Sweller et al. (2011). It must be emphasized that each of the effects described below assumes that knowledge acquired in educational institutions is domain-specific, biologically secondary information best acquired by explicit instruction. In that sense, cognitive load theory differs from most of the extant theories in the field of cognitive processes and instructional design.

The worked example effect. Learners presented with worked examples to study will perform better on subsequent problems than learners who have to solve the same problems, due to a reduction in extraneous cognitive load. Worked examples reduce working memory load compared to discovery-based problem solving and make use of the borrowing and organizing principle rather than the randomness as genesis principle. Worked examples provide explicit instruction and domain-specific knowledge.

The problem completion effect. Rather than providing a complete solution, completion problems provide a partial solution that learners must complete. Completion problems can be just as effective as worked examples and are effective for the same reasons.

The split-attention effect. Assume instructional material such as a worked example consisting of two or more sources of information that split attention and so must be mentally integrated before they can be understood. A diagram and text that are unintelligible in isolation and so must be mentally integrated provide an example.

The act of mental integration requires working memory resources that consequently are unavailable for learning, resulting in the imposition of an extraneous cognitive load. By physically integrating those sources of information, more working memory resources are available for learning, reducing extraneous cognitive load.

The modality effect. Rather than physically integrating the two sources of information as in the split-attention effect, if one source of information can be provided in spoken rather than written form, learning is facilitated. Using both visual and auditory processors rather than just the visual processor can functionally expand working memory.

The redundancy effect. Frequently, two or more sources of information can be understood in isolation. For example, text may simply repeat the information in a diagram or one source of information may in reality be uninformative and so unnecessary. Such redundant sources of information should be eliminated to reduce extraneous cognitive load, rather than integrated or converted into spoken form. The logic of the relations between the multiple sources of information is critical to determining whether information should be integrated (or presented in auditory form) or eliminated.

The expertise reversal effect. As indicated above, the storage of information in long-term memory has dramatic effects on working memory by bringing the environmental organizing and linking principle into play rather than the borrowing and reorganizing, the randomness as genesis or narrow limits of change principles. In turn, those changes necessitate changes in instructional procedures. The worked example effect provides one of many examples. As indicated above, it occurs when providing novices with worked examples facilitates learning compared to having learners solve the equivalent problems on their own. With increasing expertise in a given area of problem solving, that difference reduces and eventually reverses resulting in the expertise reversal effect. While studying a worked example may be important for a novice, it may be a redundant activity for more knowledgeable learners.

The guidance fading effect. Based on the expertise reversal effect, the explicit guidance provided by worked examples should be gradually removed as expertise increases and the environmental organizing and linking principle takes over from the other principles associated with acquiring novel information. The guidance fading effect provides evidence for this hypothesis. Only novices require explicit guidance.

The transient information effect. The introduction of modern educational technology allows a more ready use of procedures such as animations and spoken information. Sometimes, those procedures transform easily accessible, permanent information into less easily accessible, transient information. For example, transforming complex written information into spoken information can vastly increase cognitive load. Difficult to understand written information can be processed and easily re-processed on multiple occasions in a manner that is difficult or impossible with spoken information that disappears to be replaced by new information. The duration limits of working memory may render complex spoken information unintelligible. Such information is better presented in written form. Rather than facilitating learning, such technological "advances" can interfere with secondary learning.

The imagination effect. Asking learners to imagine or mentally rehearse previously learned information might assist in transferring that information to long-term memory.

The element interactivity effect. Reducing element interactivity due to extraneous cognitive load may be unnecessary if element interactivity due to intrinsic cognitive load is low. Cognitive load effects due to extraneous load should not be expected if intrinsic load is low because the number of elements that must be considered simultaneously may be within working memory limits.

The isolated elements effect. If the number of elements that must be processed is very high, it may be impossible to process them simultaneously. In that case, the information needs to be broken up into isolated elements even if that means it cannot be fully understood immediately. Understanding can come later when interacting information is reconstituted from its memorized, isolated elements.

The goal-free effect. This effect was the first cognitive load theory effect studied. Asking learners solving a mathematics problem to calculate values for as many variables as possible rather than asking them to find a value for a specific goal reduces working memory load. For example, instead of asking geometry students to "Find a value for Angle X," we can ask them to "Find the value of as many angles as you can." Attending to a specific goal may require learners to consider simultaneously the several moves needed to reach the goal. A goal-free approach limits consideration to each individual move rather than combinations of moves required to reach a goal.

Collective working memory effect. For difficult problems where knowledge is spread among two or more people, having them learn collaboratively rather than individually can facilitate learning. In effect, the group has a collective rather than an individual working memory. It should be noted that the effect disappears where all members of the group share similar knowledge.

Discussion

The architecture used by cognitive load theory with its evolutionary roots can result in instructional design recommendations that depart from many common assumptions. Nevertheless, evolutionary educational psychology provides a well-structured, highly organized base from which to consider instructional issues. All of the instructional recommendations of cognitive load theory derive from our knowledge of human cognitive architecture that was used to generate the cognitive load effects. In turn, all of those effects have been tested using multiple, replicated, randomized, controlled experiments. Those experiments provide the data that generate instructional recommendations and to the extent that those recommendations are successful provide evidence for the theory. The instructional effects discussed above can be readily understood and followed from Geary's (1995) distinction between biologically primary and secondary skills.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). Oxford, England: Academic.

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

Bruner, J. (1961). The art of discovery. *Harvard Educational Review, 31*, 21–32.

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development, 60*, 1239–1249.

Campbell, D. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review, 67*, 380–400.

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

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

Darwin, C. (2003). *The descent of man*. London, England: Gibson Square (Original work published 1871).

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

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

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

Ericsson, K. A., Krampe, R., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406.

Flynn, J. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191. doi:10.1037/0033-2909.101.2.171.

Geary, D. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist, 50*, 24–37.

Geary, D. (2005). *The origin of mind: Evolution of brain, cognition, and general intelligence*. Washington, DC: American Psychological Association.

Geary, D. (2012). Evolutionary educational psychology. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (Vol. 1, pp. 597–621). Washington, DC: American Psychological Association.

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.

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.

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

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.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

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

Popper, K. (1979). *Objective knowledge: An evolutionary approach*. Oxford, England: Clarendon.

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

Ritchie, S., Bates, T., & Deary, I. (2015). Is education associated with improvements in general cognitive ability or in specific skills? *Developmental Psychology, 51*(5), 573–82.

Siegler, R. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY: Oxford University Press.

Stotz, K., & Griffiths, P. (2004). Genes: Philosophical analyses put to the test. *History and Philosophy of the Life Sciences, 26*, 5–28.

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, NY: Springer.

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., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology, 4*, 434–458.

Thornton, A., & Raihani, N. (2008). The evolution of teaching. *Animal Behaviour, 75*, 1823–1836.

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*, 265–283. doi:10.1007/s10648-013-9243-1.