EVOLUTION OF HUMAN COGNITIVE ARCHITECTURE

John Sweller

Human cognitive architecture is peculiar. A dominant structure, working memory, is minute in its ability to process new material but massive in its ability to process very extensive and complex, previously learned information. An immeasurable quantity of that previously learned information is held in schematic form in a long-term memory that is so closely associated with working memory that it directs and, indeed, can misdirect the manner in which working memory processes information. Together, these two systems (along with a sensory memory system that is not considered in this chapter) permit us to engage in cognitive activities that can vary from simple and routine at one extreme to the intellectual heights that humans have scaled at the other extreme.

The chapter is concerned primarily with why human cognitive architecture evolved in this manner. Specifically, what are the evolutionary advantages of a working memory that requires a large long-term memory to become maximally effective in processing information but has difficulty processing new information not held in long-term memory? In answering this question, it will be suggested that working memory is very limited when handling new information because there is no central executive to coordinate novel information; working memory only becomes fully effective when handling previously learned material held in long-term memory because that previously learned material can act as a central executive; and long-term memory is very large in order to

0079-7421/03 $35.00

maximize the circumstances under which a central executive function will be available.

In order to throw light on these and other topics, the following are considered: (1) Relations between the structure of information and cognitive architecture leading to how and why some characteristic types of information impelled the evolutionary development of human cognitive architecture; (2) common information structures underlying human information processing and evolution by natural selection; and (3) consequences of the particular evolutionary directions that our cognitive architecture has taken for learning in general and for modes of presenting information.

I. How Information Structures Have Impelled the Evolution of Human Cognitive Architecture

A. Information Structures

While considerable work by many researchers over several decades has been devoted to the organization of human cognitive architecture, far less effort has gone into investigating the information structures that must have driven the evolution of that architecture. Some work has been carried out by Sweller (1994) and Halford, Wilson, and Phillips (1998). Sweller (1994) suggested that all information can be placed on a continuum according to the extent to which the elements that constitute the information interact. At one extreme, there is no interaction between the elements that need to be learned. They are independent. Element interactivity is low or, indeed, nonexistent, which means that each element can be considered and learned serially without reference to any other element. Because elements at the low element interactivity end of the continuum do not interact with each other, there is no loss of understanding despite each element being learned individually and in isolation. Understanding is defined as the ability to process all elements that necessarily interact simultaneously in working memory. Learning such material imposes a low cognitive load because each element can be learned without reference to other elements.

At the other extreme of the continuum, there is close interaction between the various elements that need to be learned. Element interactivity is high, which means that if the material is to be understood, all of the information with its multiple elements must be processed simultaneously, imposing a heavy cognitive load. Elements that interact can be processed individually, in serial fashion, but not with a high degree of understanding. Processing high element interactivity material without learning necessary relations between elements will result in rote learning. The reason rote learning occurs frequently is because learning individual elements without learning important relations and interactions between elements can reduce cognitive load dramatically. When rote learning, only one, or at most, a very limited number of elements need to be held or processed simultaneously. In effect, during rote learning, high element interactivity material is treated by the cognitive system as though it is low element interactivity material.

In contrast, learning high element interactivity material with understanding imposes very heavy cognitive demands, especially if there are many interacting elements. For understanding to occur, all interacting elements must be processed simultaneously, and for some extensive, high element interactivity material, processing all of the interacting element simultaneously may be very difficult or even impossible (Pollock, Chandler, & Sweller, 2002). Learning such material by rote reduces cognitive load, but at the cost of understanding. Examples of very low and very high element interactivity material are discussed next.

1. Low Element Interactivity Material

Laboratory-based paired associate learning tasks provide one example of learning low element interactivity material. Each paired associate can be learned without consciously considering any of the other paired associates that require learning. In that sense, the elements of the task do not interact. They can be learned in isolation without imposing a heavy cognitive load and without any loss of understanding of the task at hand.

Many realistic tasks resemble paired associate learning. Learning the names of any set of entities such as people's names, the vocabulary of a second language, or chemical symbols provide examples. Such material may be difficult to learn because there may be many elements that require learning, but the difficulty is unrelated to cognitive load. The elements can be learned in serial fashion without loss of understanding. Indeed, the concept of understanding is not normally applied to the learning of such material. One may have not learned or forgotten a particular foreign word, such as the translation of the word "cat," but one does not fail to "understand" the word. The distinction between rote learning and learning with understanding does not apply to such material. Failure to understand is reserved exclusively for high element interactivity material for which there is a heavy load if it is to be learned with understanding (Marcus, Cooper, & Sweller, 1996).

2. High Element Interactivity Material

Modern examples of high element interactivity material include learning the syntax of a second language, deriving meaning from words or symbols, balancing chemical equations, or most areas of mathematics. Examples of high element interactivity information that our ancestors had to process at a time when the human cognitive system evolved to its present point include learning a spatial layout, such as a route from point A to point B, learning to find food and shelter, learning to avoid danger, or learning complex social relations. To demonstrate the concept, the element interactivity associated with learning some of these areas is considered next.

While much of the vocabulary of a second language can be learned element by independent element with little or no interactivity, syntax cannot be learned in this manner. Elements interact and must be processed simultaneously for understanding and learning to occur. For example, word order is important in English, and word order cannot be learned without considering several words simultaneously. Consider the two sentences: "Word order is important in English" and "English in important is order word." One cannot learn that the first is grammatical but the second is not by considering each word in isolation. Learning the appropriate order of words in English requires the learner to consider all of the relevant words simultaneously. Each word and its interaction with at least some and, in some cases, all of the other words must be considered. Element interactivity is high and, as a consequence, cognitive load is high because at least at some point, all of the elements and their relations must be processed simultaneously.

Understanding and learning the structure of any mathematical process that incorporates an equation invariably involve a high degree of element interactivity. Assume a student is learning how to make a the subject of the equation $a/b = c$. In order to understand and learn the procedure, the structure of the initial equation must be considered, the numerator on the left side must be multiplied by b, which means the numerator on the right side must be multiplied by b in order to retain the equality, and the b in both the numerator and the denominator must be canceled, leaving the solution $a = cb$. While this procedure can be memorized step by step, understanding only occurs when the entire procedure can be processed simultaneously. Multiplying the left side by b without multiplying the right side by b simultaneously reflects a lack of understanding of the procedure. The entire procedure needs to be processed simultaneously if it is to be learned with understanding rather than by rote because all of the elements that need learning interact. Rote learning will reduce cognitive load substantially, but at the cost of understanding. Learning with understanding imposes a heavy cognitive load because the elements that require learning interact and so must be processed simultaneously if appropriate meaning is to be derived.

3. *An Alternative Conceptualization of Element Interactivity*

Halford, Wilson, and Phillips (1998) have provided a formal model of what they term "relational complexity" that provides an alternative to the concept of element interactivity. The model was intended primarily to provide a metric measuring individual differences, including developmental differences, in working memory. Nevertheless, it can equally provide a measure of the working memory load imposed by various tasks, especially problems that require solution. The model assumes that any task or problem can be characterized by the number of dimensions that need to be related. A unary dimension relates constants: *The cat walked*, provides an example. A binary dimension relates two variables, ternary dimension three variables, quaternary four variables, etc. The proportion $a/b = c/d$ is an example of a quaternary relation with its four variables. The number of dimensions that must be related provides the relational complexity of a task or problem, and the number of dimensions that a person can process in working memory provides a measure of working memory capacity.

Relational complexity and element interactivity may well be different terms for the same concept. Because element interactivity was devised specifically to measure differences in working memory load imposed by different tasks has been applied experimentally to a very wide variety of tasks (e.g., see Marcus et al., 1996; Sweller & Chandler, 1994; Tindall-Ford, Chandler, & Sweller, 1997) and, as shown later, has been closely related to schema-based knowledge held in long-term memory, it is used in this chapter. Nevertheless, the similarity and perhaps identity of element interactivity and relational complexity need to be kept in mind.

How has human cognitive architecture evolved to handle these information structures? In particular, how do we handle intellectually difficult, high element interactivity material? Finding one's way around new locations, understanding relations between the environment and food sources or the environment and danger, and establishing social relations and interactions with friends and enemies have been part of human life for a very long time and, along with a myriad of other activities, can all involve high element interactivity information. The nature of the mechanisms required to deal with these situations is discussed next.

B. Human Cognitive Architecture

Much more work has been carried out on human cognitive architecture than on information structures. The term "cognitive architecture" refers to the manner in which cognitive structures are organized. Cognitive structures and their relations were either discovered or emphasized as individual structures by various researchers since the early 1930s and have been conceptualized into a unified architecture since the early 1970s. While there are many active research areas and controversies associated with that architecture, there is also a substantial degree of consensus concerning its basic outline. This section describes those aspects of human cognitive architecture around which there is broad agreement, including a brief history of our developing understanding of the topic.

1. *Working Memory*

Initially designated short-term memory (e.g., Miller, 1956), it is now more commonly referred to as working memory (e.g., Baddeley & Hitch, 1974) to reflect the change in emphasis from a holding store to the processing engine of the cognitive system. Working memory is the seat of consciousness and, indeed, can be equated with consciousness in that the characteristics of our conscious lives are the characteristics of working memory. The most commonly expressed attributes of working memory are its extremely limited capacity, discussed by Miller (1956), and its extremely limited duration, discussed by Peterson and Peterson (1959). In fact, both of these limitations apply only to novel information that needs to be processed in a novel way. Well-learned material, held in long-term memory suffers from neither of these limitations when brought into working memory (Ericsson & Kintsch, 1995).

While initially conceptualized as a unitary concept, working memory is now more commonly assumed to consist of multiple streams, channels, or processors. For example, Baddeley (e.g., Baddeley, 1992; Baddeley & Hitch, 1974) divided working memory into a visuospatial sketch pad for dealing with two-dimensional diagrams or three-dimensional information, a phonological loop for dealing with verbal information, and a central executive as a coordinating processor.

A major consequence of the limitations of working memory is that when faced with new, high element interactivity material, we cannot process it adequately. We invariably fail to understand new material if it is sufficiently complex. In order to understand such material, other structures and other mechanisms must be used. Processing high element interactivity material requires the use of long-term memory and learning mechanisms.

2. *Long-Term Memory*

Because we are not conscious of the contents of long-term memory except when they are brought into working memory, the importance of this store and the extent to which it dominates our cognitive activity tend to be hidden from us. Given this hidden nature of long-term memory, it is not surprising that modern research into long-term memory postdated research into working memory. It took some time for researchers to realize that long-term memory is not just used to recognize or recall information but rather is an integral component of all cognitive activity, including activities such as high-level problem solving. When solving a problem, it was previously assumed that knowledge stored in long-term memory was of peripheral rather than central importance. De Groot's (1965) work on chess (first published in 1946) demonstrated the critical importance of long-term memory to higher cognitive functioning. He demonstrated that memory of board configurations taken from real games was critical to the performance of chess masters. The significance of this finding became fully apparent with Chase and Simon's (1973) paper on the same topic.

3. Schemas

Knowledge is stored in long-term memory in schematic form, and schema theory describes a major learning mechanism. Schemas allow elements of information to be categorized according to the manner in which they will be used. Thus, for example, we have a schema for the letter *a* that allows us to treat each of the infinite number of printed and hand-written variants of the letter in an identical fashion. Schemas first became important cognitive constructs following the work of Piaget (1928) and Bartlett (1932). They became central to modern cognitive theory, especially theories of problem solving, in the 1980s. As well as the work of de Groot (1965) and Chase and Simon (1973), Gick and Holyoak (1980, 1983) demonstrated the importance of schemas during general problem solving, and Larkin, McDermott, Simon, and Simon (1980) and Chi, Glaser, and Rees (1982) demonstrated the critical role of schemas in expert problem solving. As a consequence of this work, most researchers now accept that problem-solving expertise in complex areas demands the acquisition of tens of thousands of domain-specific schemas. These schemas allow expert problem solvers to recognize problem states according to the appropriate moves associated with them. Schema theory assumes that skill in any area is dependent on the acquisition of specific schemas stored in long-term memory.

Schemas, stored in long-term memory, permit the processing of high element interactivity material in working memory by permitting working memory to treat the many interacting elements as a single element. In effect, the interacting elements are buried within the schema that, as discussed in more detail later, can act as a central executive by appropriately coordinating those interacting elements. As an example, anyone reading this chapter has schemas for the complex squiggles that represent a word. Those schemas, stored in long-term memory, allow us to reproduce and manipulate the squiggles that constitute writing, in working memory, without strain. However, we are only able to do so after several years of learning.

4. *Automation*

Everything that is learned can, with practice, become automated. After practice, specific categories of information can be processed with decreasing conscious effort. In other words, processing can occur with decreasing working memory load. As an example, schemas that permit us to read letters and words must initially be processed consciously in working memory. With practice they can be processed with decreasing conscious effort until eventually, reading individual letters and words becomes an unconscious activity that does not require working memory capacity. Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977) demonstrated the contrast between conscious and automated processing. In his versions of the ACT architecture, Anderson places a heavy emphasis on automation (e.g., Anderson & Lebiere, 1998). Kotovsky, Hayes, and Simon (1985) demonstrated the enormous benefits of automated processing to problem-solving skill. A problem isomorph that could be solved using automated rules was solved 16 times more rapidly than an isomorph that required the rules to be processed consciously. Thus, high element interactivity material that has been incorporated into an automated schema after extensive learning episodes can be manipulated easily in working memory to solve problems and engage in other complex activities.

5. *Coalescing of Isolated Cognitive Structures and Functions into a Unified Cognitive Architecture*

While these cognitive structures and functions are studied frequently in isolation, they can be combined into a unified cognitive architecture. Atkinson and Shiffrin (1968) elucidated relations between working or short-term memory and long-term memory. In depicting the flow of information between memory stores, they presented a cognitive architecture that is at the core of most subsequent treatments. The cognitive architecture described here incorporates the Atkinson and Shiffrin (1968) model along with the two learning mechanisms, schema acquisition and automation.

All conscious processing of information consists of the manipulation of schemas, which can act as interacting elements, in working memory. That manipulation can result in learning, which consists of the creation of new, higher order schemas and automation. Schemas are stored in long-term memory. They can only be brought into working memory if they are held in long-term memory. The primary, possibly sole, function of long-term memory is to hold hierarchically organized schemas. The limitations of working memory refer to its limited ability to process separate schemas that have not been incorporated into a higher level schema. Only a very small number of schemas can be processed and they can only be held in working memory for a few seconds. Some schemas can consist of a huge number of interacting elements. These interacting elements are lower level schemas. When brought into working memory, a schema, no matter what its size, is treated as a single element. Thus, schemas have a dual function of organizing information in long-term memory and reducing working memory load. Automation has a similar function of reducing working memory load. On this analysis, the two learning mechanisms of schema acquisition and automation both have a primary function of reducing working memory load and so allowing a limited working memory to process large amounts of information, providing that information has, after learning, been stored in long-term memory in the form of automated schemas. This configuration of cognitive structures and functions has evolved to handle the information humans must deal with.

C. Coordination of Information Structures and Cognitive Architecture

The information structures and cognitive architecture described in the previous sections can be assumed to be closely coordinated. Biological evolution could be expected to ensure that coordination. The particular information structures that the cognitive configuration has to deal with can be expected to have been a major governing factor in the direction of the evolution of that configuration. Accordingly, it is appropriate to establish links between information structures and cognitive structures and, in the process, attempt to answer questions concerning aspects of human cognitive architecture. An important consideration is how human cognitive architecture evolved to deal with high element interactivity material.

1. Schemas, Working Memory, and High Element Interactivity Material

High element interactivity material, by its very nature, must be processed simultaneously in working memory. It cannot be processed element by individual element and still retain its meaning. One might assume that the obvious way human cognitive architecture would evolve to handle such material would be to develop a sufficiently large working memory to handle many interacting elements simultaneously. Our cognitive architecture did not, of course, follow this route. For reasons discussed later, humans have not developed a large working memory when dealing with new information. As a consequence of our limited capacity working memory, we are not able to process novel, high element interactivity material. When faced with novel information that contains many interacting elements, we inevitably fail to understand it. Understanding requires all interacting elements to be processed simultaneously, at least at some point, and when confronted with many interacting elements, processing all of them simultaneously in working memory is impossible. As indicated earlier, if we feel impelled or motivated to process such information, the best that can be done is to rote learn some aspects of the material.

Rather than develop a large working memory to handle novel, information-rich, high element interactivity material, our cognitive architecture has evolved to deal with such information by first integrating it into schemas held in long-term memory. Interacting elements can be incorporated within a schema and that schema can then be treated as a single element within working memory. Because those schemas can be processed in working memory as a single element, they eliminate the problem of a limited working memory. Our cognitive architecture has evolved so that very high element interactivity material encompassing large amounts of information can *only* be handled when incorporated in schemas. It follows that such material can only be fully processed in working memory after extensive learning has occurred, sometimes over very long periods of time. Until learning through schema acquisition and automation has taken place, the human cognitive system cannot adequately deal with very complex, high element interactivity material. After learning, such information rich material is handled easily and smoothly.

2. *When Working Memory Is Unlimited*

A limited capacity working memory is a central concept in cognitive psychology. Since Miller (1956) and Atkinson and Shiffrin (1968), most discussions of human cognitive architecture have incorporated a limited capacity short-term or working memory. It is appropriate that they should do so because working memory is limited when dealing with new information. Nevertheless, capacity limitations only apply when dealing with new, not old information. When dealing with previously learned material, the only discernible limit on working memory is the amount that has been learned and stored in long-term memory. Massive, seemingly unlimited amounts of information can be processed by working memory providing they have previously been incorporated in schemas. A schema may contain a large amount of information but will be processed in working memory as a single element.

The tension between a very limited working memory when dealing with new information and an unlimited working memory when dealing with learned material can be seen as far back as Miller's (1956) paper. Miller's concept of "chunking," which today can be incorporated in the more sophisticated conception of schema construction, altered the amount of information that short-term memory could hold. By chunking together elements of information, the amount of information held by short-term memory could be increased. In that sense, learning could be used to increase the effective capacity of short-term memory. Similarly, while working memory can only process a limited number of schematically based elements, what constitutes an element is entirely dependent on what has been learned. If much has been learned, an element can incorporate a massive amount of information. Indeed, there may be no limit to the amount of information incorporated in a schema that acts as a single element in working memory. In that sense, there is no limit to the amount of information that can be processed by working memory. The capacity limitations of working memory appear only when new, unorganized information that has not yet been organized into schemas must be processed.

Empirically, de Groot (1965) and Chase and Simon (1973) provided the strongest early evidence for this phenomenon. Chess experts with their appropriate schemas can hold an entire board of chess pieces taken from a real game in working memory because they have a schema for that configuration. Novices have to remember each piece individually, which is beyond working memory capacity, as are random configurations for experts. This result has been obtained in a wide variety of areas (e.g., Egan & Schwartz, 1979; Jeffries, Turner, Polson, & Atwood, 1981; Sweller & Cooper, 1985).

The ability of working memory to hold and process large amounts of learned information for long periods of time was recognized by Ericsson and Kintsch (1995). Their concept of "long-term working memory" applies to very well-learned material. For such material, the capacity limitations of "short-term working memory" disappear. Large amounts of domain-specific, well-learned material in complex areas such as text comprehension, chess, and music can be held and processed in working memory for long periods. The usual capacity and duration limits associated with working memory are not in evidence for such well-learned material.

In effect, we are dealing with two continua: A learning continuum and a working memory limitations continuum. At one extreme, when dealing with yet-to-be-learned or unlearned material, well-known working memory limitations are relevant to processing. At the other extreme, when dealing with well-learned material, the usual working memory limitations are irrelevant and working memory can best be described in terms of Ericsson and Kintsch's (1995) long-term working memory. Thus, in this chapter, long-term working memory is incorporated at one end of a working memory continuum rather than as a discrete structure. Rows 1 and 5 of a cognitive

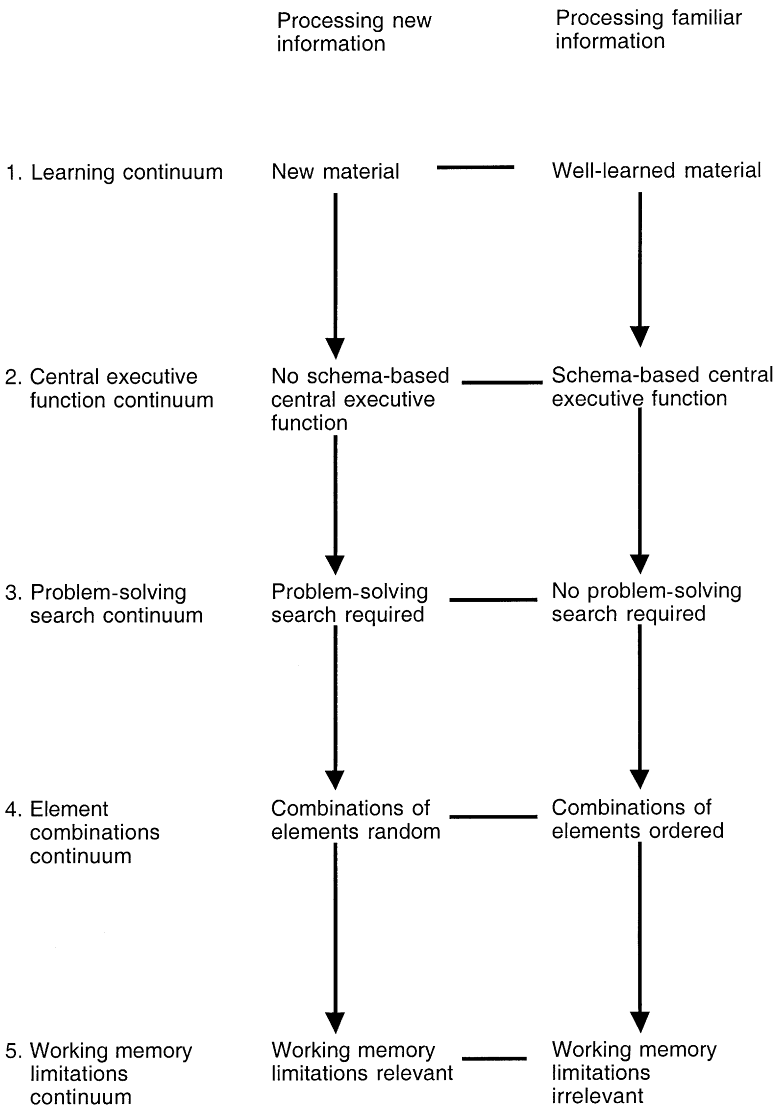

Fig. 1. A cognitive matrix of continua.

matrix of continua (see Fig. 1) depict the learning and working memory continua, respectively.

This chapter is concerned primarily with the intervening constructs relating unlearned material and a limited working memory at one extreme of the matrix of continua and relating well-learned material and an unlimited working memory at the other extreme. While learning through schema acquisition eliminates the problem of a limited working memory having to handle large numbers of interacting elements (the right side of the matrix of continua), the question remains why human cognitive architecture evolved in this manner rather than following the apparently more straightforward route of a larger working memory, either as an adjunct to or even as a substitute for some learning mechanisms. That route of a larger working memory would have permitted larger amounts of new material to be processed. There should be evolutionary reasons why that route was not followed.

D. A Cognitive Matrix of Continua

Information we deal with can be placed on a learning continuum extending from new material for which there are very limited schemas available to well-learned material with its elements incorporated into an extensive schematic framework. The first row of Fig. 1 indicates the two extremes of this learning continuum.

The second row is concerned with schemas. While the characteristics and functions of schemas were discussed previously, they have one additional function that is less commonly considered: Schemas held in long-term memory provide working memory with a central executive. Furthermore, they may be the only structure available to provide a central executive for working memory. The second row of Fig. 2 indicates the two extremes of the schema-based, central executive function continuum.

A schema, acting as a central executive, coordinates information. It indicates which information can be ignored, which information is significant, and how the elements of significant information relate to each other. A well-established, automated schema acts exactly as we would expect an effective central executive to act. Both incoming information and the responses to that information can be governed and coordinated by schemas. Provided schemas are available, no other central executive function is required for humans to process information. Of course, schemas must be learned and activated and so are not always available.

Evidence for the central executive function of schemas comes from one of the conditions under which problem solving fails. If a problem solver learns to solve a class of problems using a particular technique, he or she will continue to attempt to use the technique even when presented a problem with a similar surface structure for which it is inappropriate. This mental set, or *Einstellung*, was demonstrated by Luchins (1942) using his well-known water jar problems (see also Ben-Zeev & Star, 2001; Fingerman

& Levine, 1974; Levine, 1971; Ross & Kilbane, 1997; Sweller, 1980a,b; Sweller & Gee, 1978.) The effect occurs because a schema is acquired when learning to solve an initial set of structurally similar problems. That schema then directs the solution of all subsequent similar problems in exactly the manner to be expected of a central executive. On the one hand, it permits the solution of problems that would be quite insoluble without an appropriate schema. On the other hand, it continues to organize the elements and solution procedures of other, structurally dissimilar problems that have similar surface features even when the solution procedures are quite inappropriate. As a consequence, the solution will either be delayed or fail entirely. In contrast, a person presented such a target problem without first having acquired the inappropriate schemas will have no difficulty solving it. The frequently spectacular contrast between the performance of people with and without inappropriate problem solving schemas demonstrates Einstellung. In the process, the central executive function of schemas is revealed graphically.

While schemas held in long-term memory provide a central executive for working memory at the well-learned end of the learning continuum, it can be argued that there is no available central executive at the other end of the continuum when dealing with new, yet-to-be-learned material. Two arguments can be put forward against the notion of a coordinating central executive when dealing with new, yet-to-be-learned information for which no schema is available. The weaker argument simply states that the characteristics of a central executive have not been sufficiently well specified to be assured that it exists and, in any case, there is no real empirical evidence for any possible central executive-type construct. This argument is not pursued further because it is overridden by the stronger argument, which is that the very concept of a central executive dealing with yet-to-be-learned material in a nonrandom manner leads to an infinite regress and so is logically impossible.

Consider a central executive coordinating new information in a nonrandom manner. The executive must make decisions on how information is to be dealt with in that it must decide which elements will be combined, coordinated, or related in some fashion. In other words, it must decide on how the information will be processed. That information is both new and infinite in scope. It is new in the sense that the executive has not dealt with such information before and it is infinite in that there is no limit on the types of information or how that information will have to be combined or processed. Other than randomly, how does the central executive decide how to deal with this potentially infinite range of new information? It cannot draw on previous knowledge because the material is new. It could use biologically programmed or “hardwired” procedures for a limited number of activities but not for the infinite range of information that humans can potentially deal with. (It will be assumed that we are not hardwired to deal with each of the procedures of complex mathematics, for example.) If these assumptions are correct, there is only one other way a nonrandom central executive can arrive at a decision. If the information is to be dealt with in an orderly fashion, it must have another executive function available to direct it. However, the logic of a second executive will, of course, be identical to the logic of the first, requiring a third executive, etc. This infinite regress indicates that the entire concept is flawed and requires replacing. Mechanisms other than a schema-based central executive are required to coordinate new, unlearned information.

If there is no central executive available to coordinate new, yet-to-be-learned elements, how are these elements dealt with? Research into problem solving provides an answer and also provides the third row, the problem-solving search continuum of the matrix of continua. Problem solving search is required precisely when we are faced with new information for which we have yet to acquire appropriate schemas. Critical research in the early 1980s on expert–novice distinctions (e.g., Chi et al., 1982; Larkin et al., 1980) clearly established that when faced with a novel problem for which a learned solution is not available (i.e., a problem being dealt with by a novice with respect to that class of problems), people engage in problem-solving search using a weak strategy such as means-ends analysis (Newell & Simon, 1972). Using this strategy, problem-solving moves are generated by attempting to find operators that will reduce differences between each problem state attained and the goal or a subgoal. In other words, faced with a novel situation, people use general problem-solving search strategies in an attempt to impose some order and choose between various element combinations. The purpose of those search strategies is to attempt to coordinate yet-to-be-learned elements with the external environment. This process of matching is only required when faced with new material for which adequate schemas have yet to be acquired. With respect to the cognitive matrix of continua of Fig. 1, problem-solving search flows directly from the left side of the first two rows of the matrix of continua. That is, it occurs because a person is dealing with new, unlearned material for which there is no schema-based central executive.

At the well-learned end of the continuum, problem-solving search is unnecessary. On the right side of the matrix, when dealing with well-learned material for which well-established schemas are available, the schemas themselves generate problem-solving moves (Larkin et al., 1980). Problem-solving search to coordinate and establish relations between elements is unnecessary because schemas provide all of the necessary relations. In between the two extremes of the third row of the matrix, search becomes less and less important, moving from the point where moves are generated by problem-solving search to the point where they are generated by schemas. Thus, the third continuum, the problem-solving continuum, has been established and related to the learning and central executive continua.

The first three continua lead to the critical fourth continuum that provides a direct explanation for working memory characteristics when dealing with both new and well-learned material. On the left side of the matrix, operators and problem states must be chosen during problem-solving search in the absence of schemas and their executive function. A major function of problem-solving search is to impose a degree of order on otherwise disordered, more or less random, combinations of elements. This order is imposed by attempting, as far as possible, to use the environment to provide appropriate relations between elements. Random combinations of elements are held in working memory, and attempts are made by problem-solving search to order them in a manner that reflects the environment. Once an appropriate set of relations has been established, the goal of the problem has been attained.

It is frequently forgotten that by necessity, problem-solving search conducted without solution knowledge of moves or element combinations must include a random component. Consider means-ends analysis as an example of a strategy that does not rely on a heavy knowledge base. This strategy requires considerable control and has a relatively small random component. Nevertheless, a random component cannot be fully eliminated. The strategy involves first choosing a move and then testing it to see whether it reduces differences between a current problem state and the goal or a subgoal state. Checking whether a move reduces differences between the current problem state and the goal state cannot occur before the move has been chosen. It must occur after the move has been chosen. If there is no prior knowledge concerning the effect of the move (in the form of schemas or partial schemas), it must be chosen randomly. Only after it has been chosen can it be assessed for effectiveness. There is a high degree of control in that differences between current and goal states are extracted before moves are chosen and moves that do not reduce differences between the current and goal states are rejected. Nevertheless, in the absence of prior knowledge, which moves are chosen for testing using the means-ends heuristic must be random. In the absence of a central executive, there is no other technique available. Other than a random mechanism, there can be no knowledge-free procedure for initially choosing moves to test to see if they reduce differences between current and goal states. As a consequence, on the left extreme of the element combinations continuum, random combinations of elements are necessarily the norm.

With random choice, the greater the number of alternative subgoals and operators from which to choose while problem solving, the less likelihood that an appropriate choice will be made. As the number of choices available increases, the probability of a choice leading to a dead end also increases. With increased choice, problem-solving search becomes decreasingly effective and, indeed, with even a moderately large number of choices, search becomes pointless. Making an appropriate choice out of two or three at each choice point is feasible. Choosing out of several dozen or more alternatives at each choice point would render the process futile. Problem-solving search is more likely to be effective if it can be limited, and our cognitive architecture had to evolve to ensure that it is always limited because anything beyond a small search space reduces the probability of arriving at a solution to almost zero.

With increasing knowledge, the random choice of elements decreases. At the right extreme of the element combinations continuum, well-learned material has schemas to coordinate elements, and problem-solving search is unnecessary with all element combinations ordered by previously acquired schemas. It is only after learning has occurred that problem-solving search is not needed to order elements because they are ordered by schemas.

We are now in a position to consider the last continuum, the working memory limitations continuum (the fifth row of the cognitive matrix of continua), and to indicate why working memory must be limited when dealing with new, yet-to-be-learned material. The need for a random component when combining elements through problem-solving search leads directly to a requirement for working memory to have a severely limited capacity. Consider someone dealing with two new elements. While the manner in which elements should be combined will vary depending on the material being dealt with, assume that they must be combined using the logic of permutations. There are two (2!) unique ordered permutations possible for two elements (*ab* or *ba*). As the number of elements increases, the number of permutations rapidly becomes very large ($5! = 120$). The way in which these elements should be combined can be handled easily by a system with a schema-based central executive determining the appropriate combination, as occurs on the right side of the matrix of continua, dealing with well-learned material. Without a central executive, on the left side of the matrix dealing with new material requiring problem-solving search and its attendant need to combine elements randomly, no more than two or three elements can be handled because any more elements would result in more potential combinations than could be tested realistically.

It may be for this reason that we have evolved with a limited working memory. When dealing with new, interacting elements that have not been learned (i.e., have not been formed into schemas), there is no structure that can indicate the manner in which the elements should be combined and so the need to combine any more than two or three elements can result in a huge number of possible combinations that could not be tested properly against reality. A limited working memory ensures that combining a large number of elements in the absence of a controlling schema cannot occur. Such combinations of many elements would rarely reflect reality. The proposal that working memory is limited in order to limit the number of element combinations that require testing constitutes a central core of this chapter.

The suggestion that a limited working memory may have advantages when processing information under some conditions has been made previously. Both Dirlam (1972) and MacGregor (1987) provided a formal analysis indicating that search is most efficient when the number of items being dealt with closely approximates the number of items that can be held in working memory. Elman (1993) and Newport (1990) suggested that by constraining the search space for grammatical forms, a limited working memory is an advantage when learning a language. Kareev (1995) indicated that when dealing with correlations, a smaller sampling size increases the probability of the sample having a correlation stronger than the population. Thus, if a relation exists, a limited capacity working memory could have the effect of increasing the probability of its being detected. Kareev, Lieberman, and Lev (1997) provided data indicating that people with smaller working memories were more likely to perceive a correlation than people with larger working memories. Taken together, these suggestions all indicate that there may be advantages to a limited working memory when dealing with new material, and the commonsense view that a larger working memory should be advantageous may be erroneous.

In summary, the manner in which our cognitive architecture interacts with information can be represented by a matrix that incorporates five parallel continua: (1) a yet-to-be-learned to well-learned continuum in which the extent that individuals have learned the material (i.e. acquired schemas) that they are faced with increases; (2) an uncontrolled to schematically controlled central executive function continuum in which the degree to which schemas control working memory processing increases; (3) a problem-solving search continuum in which the need to solve problems by problem-solving search varies from essential to unnecessary; (4) a random to ordered combination of elements continuum in which the manner in which elements combine varies from random to ordered; and (5) a working memory limitations continuum with working memory limitations critical at one end and irrelevant at the other.

These five continua are linked causally providing a matrix. On the left side of the matrix, new material that is still to be learned has no central executive coordinating high interactivity elements. Some degree of coordination only can be provided by problem-solving search that incorporates testing the effectiveness of random combinations of elements. When dealing with these element combinations, a limited capacity working memory is essential to prevent a combinatorial explosion. In contrast, on the right side of the matrix, well-learned material has schemas providing a central executive function. Problem-solving search is not required because schemas provide ordered combinations of elements. Interacting elements are incorporated within schemas, resulting in no effective working memory limits when dealing with such well-learned material. Examples demonstrating the relations incorporated in the matrix are discussed in detail in the next two sections.

1. Processing Well-Learned Material

Assume a person is faced with a high element interactivity task such as navigating from one location to another in a city. How the person deals with that task depends on the learning continuum. The right side of the matrix of continua of Fig. 1 is considered in this section. At this extreme, the person will have learned all that is needed to handle the information using automated schemas. Where to turn, the consequences of being in one traffic lane rather than another, and even where there are bumps or potholes in the road are all incorporated in appropriate schemas. At this extreme of the matrix of continua, schemas act as a central executive when brought into working memory. They coordinate the huge number of sensory inputs and motor outputs with virtually no load on working memory. All the myriad of elements associated with driving from point *a* to point *b* are ordered and organized by the appropriate schemas. The driver will not engage in problem-solving search and may arrive at the destination with almost no conscious effort. Working memory limitations do not impinge on performance at this end of the continuum because the automated schemas generating actions do not impose an appreciable working memory load. Other activities requiring working memory, such as holding a conversation or thinking about an unrelated activity, can be carried out easily because little working memory capacity is required for navigation.

Similarly, for any well-learned activity, such as reading a book, using a computer, going for a walk, and listening to music, schemas tell us what to listen or look at, what to do, and when to do it. For such material, the well-learned nature of the information permits schemas to govern and coordinate the various elements; this central executive function of schemas allows huge amounts of information to be both held and processed in working memory. Problem-solving search to establish appropriate relations between elements does not occur. It has no function because suitable schemas determine all relations between elements. Under these conditions, working memory limitations are not in evidence (Ericsson & Kintsch, 1995), providing we do not come across new, unfamiliar material for which we have not acquired schemas. When faced with new, unlearned material (i.e., material for which a schema is not available to act as a central executive) different processes are required.

2. *Processing Novel, Yet-to-be-Learned Material*

In contrast to a traveler at the highly learned end of the learning continuum, consider someone at the other end of the continuum, represented by the left side of the matrix of continua of Fig. 1. This person is traversing the route for the first time and so has few or no schemas to coordinate the elements of information. There is no well-defined, schema-based central executive to deal with the information. In the complete absence of a schema-based central executive, problem-solving search to ascertain a suitable route will be required. As indicated earlier, when engaged in problem-solving search, at certain points there is no choice but to combine and test elements randomly. In this particular case, that requires choosing roads on a random basis and testing the consequences of the choice either mentally or physically. That means while we can consider the consequences of choosing a particular direction, we can only do so after deciding to consider that direction, not before. In the absence of knowledge, the decision to choose a particular direction for consideration must be random. More frequently, partial executive functions can be provided by other sources (e.g., a map) and, indeed, precise, ongoing instructions from someone else can provide full executive functions. Nevertheless, in the absence of suitable domain-specific schemas to coordinate elements of information, the person normally will need to engage in problem-solving search using a general problem-solving strategy such as means-ends analysis. Using this problem-solving strategy, the problem solver must attempt to find problem-solving operators that will reduce the differences between the current problem state and a goal or subgoal state. These operators must be chosen randomly but can be tested mentally for their consequences using means-ends analysis, a process that is very expensive in terms of the limited working memory resources available at this extreme of the matrix of continua (Sweller, 1988).

The left side of the matrix of continua applies to a wide variety of intellectual tasks. When listening to or reading unfamiliar, high element interactivity material, various aspects of the material need to be related in order to derive meaning. If the relations are not incorporated in schemas, they will need to be processed in working memory, which will require a problem-solving process to determine which relations are appropriate. Initial attempts to establish connections between referents, for example, will contain random components and so some attempted relations will be inappropriate and fail, resulting in a comprehension failure. To understand the statement "Initial attempts to establish connections between referents will contain random components," the listener or reader must establish that "random components" refer to the "attempts" and not the "connections" or "referents" directly. To understand text, the number of such attempted relations must be limited in order to prevent a numerical explosion of possible relations that would permanently prevent comprehension. A limited working memory reduces the number of possible relations allowing the prospect of comprehension. Nevertheless, if there are too many possible relations not previously incorporated in schemas, comprehension will fail (e.g., Britton & Gulgoz, 1991). In contrast, schematic control determines which relations between interacting elements are appropriate and embeds them within schemas. A schema for a statement includes all of the interacting elements within it and can be processed readily in working memory. As a consequence, large amounts of information can be processed with a limited working memory load, allowing very complex relations to exist and thus ensuring comprehension.

II. Human Information Processing Recapitulates Evolution by Natural Selection

The manner in which information is processed by the human cognitive system, as described earlier, recapitulates the manner in which natural selection handles information of the genetic code that results in the perpetuation and evolution of species. Both systems consist of very large bodies of information that control the activities of natural entities that must continually adapt their behavior to a complex environment. It can be argued that the structure of such information systems happens to have certain fixed characteristics irrespective of the particular entity they control or the specific activities of that entity. As a consequence, both natural selection controlling the adaptation of organisms to their environment and the cognitive structures that control human behavior incorporate a single, natural system of information that underlies both processes.

There are several features of such a natural system of information. (1) Natural information systems consist of an information store sufficiently massive to permit them to behave appropriately in a complex environment. (2) Any alteration or variation to the information store is tested against the environment for effectiveness with effective alterations added to the store while ineffective alterations are deleted. (3) All natural variations to the store are necessarily random. (4) Because large random variations will almost certainly destroy the functionality of the store, mechanisms must exist to ensure that most variations are small. The validity of each of these propositions is considered in more detail.

A. The Size of Information Stores

Information stores that coordinate activity with a complex, natural environment over extended periods of time are necessarily massive. Many natural environments are complex in the sense that they can be characterized by a large variety of states. While any single, simple physical attribute of an environment, such as temperature, pressure, radiation, or chemical composition, may have narrow limits under some circumstances, combinations of attributes frequently result in a constantly altering environment. Information stores governing the activity of an entity must be capable of coordinating that activity with its variable environment. In general, the more variable an environment, the greater the size of the information store required to coordinate activity with that environment. The complexity of an environment must be matched by a commensurately complex information store.

The genome of a species provides an example of the required size of a natural information store. The genetic information contained within the genomes of organisms surviving in complex environments must be massive in order to permit survival. The human genome consists of about 3 billion base pairs. While much of this information appears not to be used in genes, humans still have an estimated 30,000 or more genes. This enormous store of information is required to coordinate complex human activity with our environment. In contrast, the much simpler activity of yeast requires about 1/200th of the number of base pairs and approximately 1/5th of the number of genes of a human. The simpler activity of yeast requires a much smaller store of information. Nevertheless, in an absolute sense, even information stored in the genome of yeast is very large. (It also needs to be noted that there may be no simple numerical contrast that can be used to correlate genetic factors and species complexity. While there may be some correlation between the number of base pairs in the DNA of species and their complexity, some very simple species have many more base pairs than humans. Furthermore, the recent consensus that humans have about 100,000 genes has been broken since the successful mapping of the human genome. The estimated number of genes now varies from 30,000 to 40,000 with the lower number more probable. That number is only marginally larger than for a plant. Complexity may be incorporated in each gene rather than expressed by the number of genes. It appears that human genes are more complex than that of simpler organisms, with human genes generating more protein products. See Aparicio, 2000; International Human Genome Sequencing Consortium, 2001.)

The large store of information contained within a species' genome is mirrored by the large store of information held in human long-term memory. Information held in long-term memory governs human behavior in an analogous manner to a genetic code governing the behavior of a species. Rows 1 and 2 (the learning and central executive function continua) of the cognitive matrix of continua depicted in Fig. 1 can be used to substantiate the analogy. On the right side, a very large store of well-learned material determines much human behavior. Similarly, a large store of genetic information determines the characteristics of a species. Human behavior is not permanently fixed, and the left side of the learning and central executive continua reflects the fact that common patterns of behavior must alter to reflect a changing environment. Because genetic characteristics of a species must also change to reflect a changing environment, mechanisms to affect genetic change are built into the genetic system.

B. Testing the Effectiveness of Variations in an Information Store against an Environment

The manner in which variations to natural information stores are tested for effectiveness can be described by rules. The general rule is that a variation that more closely coordinates activity with an environment will tend to persist, whereas a variation that decreases the coordination of activity with an environment will disappear. This rule is referred to as the environmental coordination rule. Particular versions of this general rule can be described for both evolutionary biology and the manner in which human cognitive architecture handles information.

The mechanism of natural selection is well known. Offspring retain many of the characteristics of their parents, and individuals with more advantageous variations leave more offspring than individuals with less advantageous variations. Natural selection is an example of the environmental coordination rule. Information contained in a genetic code will persist if that code results in activity that is well coordinated with an environment. Information will disappear if activity is poorly coordinated with an environment. An alteration that increases coordination of activity with an environment will result in permanent changes to the genetic code. An alteration that decreases coordination with an environment will result in no permanent changes to the genetic code.

The environmental coordination rule applies equally to humans processing information. The rule is reflected in the third row of the matrix of continua, the problem-solving search continuum. Humans will generally use information in long-term memory to govern their activity (on the right of the problem-solving continuum). Any departures from the use of that information will be tested for effectiveness against the environment using problem-solving strategies such as means-ends analysis. Novel procedures that coordinate activity with the environment more accurately are likely to be retained in long-term memory and used again. The long-term memory store is altered by successful procedures. Procedures that fail to coordinate with the environment will not be retained in long-term memory and tend not to be used again. The long-term memory store is left largely unchanged by unsuccessful procedures. This mechanism is closely analogous to evolution by natural selection.

C. Random Variations to Natural Information Stores

Variations to natural information stores occur randomly. Random genetic variation mechanisms are well known. Mutation and sexual recombination result in random variations and without these mechanisms, no natural alterations to a genetic code would occur. Barring deliberate human action, there is no other mechanism available. Similarly, and as indicated earlier, barring knowledge held in long-term memory indicating which moves to make when faced with a problem, moves can only be generated randomly as indicated on the left side of the elements combinations continuum of the matrix of continua. Until the knowledge base can be brought into play allowing movement to the right side of the elements combinations continuum, move generation is necessarily random, just as mutation and genetic recombination are random. Material deliberately intended to have an educative function provides the only exception to these mechanisms. Education techniques can reduce or eliminate the random generation of problem-solving moves (see later), just as the deliberate alteration of a genetic code substitutes for the random variations due to mutation and genetic recombination.

Both the historical reasons for and the consequences of the concept of random variations to natural information stores need to be carefully noted. Random variation was required to explain the evolution of species through natural selection without a guiding intelligence and provides one of the major functions of the theory of evolution. In other words, evolution by natural selection does not have a "central executive" to guide the process. Indeed, in the many theologically motivated debates concerning the theory of evolution, there appears to have been a tacit consensus that no natural, as opposed to supernatural, candidate for an intelligence guiding the evolution of species was available. All of the "intelligence" of the system resides in genes. A requirement for a second intelligence to guide the manner in which genes evolve would require a third to guide the second and so on, resulting in an infinite regress. Random mutation and natural selection act as substitutes for an additional intelligence.

One purpose of this chapter is to suggest that human cognitive architecture similarly has no natural intelligence in the form of a central executive guiding the generation of novel procedures. There is a natural intelligence in the form of schemas held in long-term memory that guide previously learned procedures that have been established as effective. Those schemas govern the vast bulk of human behavior, including determining what new material should and should not be learned. As indicated previously and as is the case for evolution by natural selection, that stored information incorporates intelligence. An additional intelligence (or central executive) would require an infinite regress to function. When schematic knowledge held in long-term memory is not available or when guidance from other humans who hold such knowledge is not available, only random selection of mental actions is possible. Of course, knowledge gained from those randomly selected mental actions can be retained in long-term memory, which ensures that subsequent actions are intelligent rather than random. Analogously, genetic codes provide a natural intelligence to guide the continuation of a successful species. When suitable codes are not available, random mutations determine which codes will be offered to the environment for testing as part of the processes of natural selection.

D. The Size of Random Variations to a Natural Information Store

Natural information stores have mechanisms to ensure that variations to the store are small. If, in order to deal with a very complex, variable environment, a store is very large, then relative to its size, any usable alterations will constitute a minute proportion of the total store. A large variation in the store will almost certainly disrupt essential functions and so is incompatible with the continuation of a natural store in a natural environment.

Individual mutations and genetic recombination that permit continuation of a species constitute a very small proportion of a genetic code. A substantial genetic shift will take many thousands or even millions of years. The huge overlap in the genetic code of species that separated millions of years ago is a testament to the stability of genetic codes. Changes over short periods are minute. Only such small variations are viable. Large variations do not survive. Similarly, as indicated by the working memory limitations continuum, human working memory ensures that alterations to the long-term memory store are relatively slow and small.

In summary, mutation and sexual recombination result in quite random variations analogous to the random choice of moves faced by a person solving a problem for which schema-based solutions are not available. The usefulness or otherwise of a genetic variation can only be assessed after it has occurred. If it is successful, information in the genetic code will be passed on to subsequent generations, whereas a failure will result in a genetic dead end with the information not passed to subsequent generations. Similarly, when limited or no knowledge is available to a problem solver, moves must be chosen randomly. Successful moves may be incorporated in schemas that then can be used indefinitely when faced with similar circumstances. Unsuccessful moves result in dead ends with information not incorporated in schemas and not used subsequently.

Under this formula, a schema encapsulates psychological information in the same way that a gene encapsulates genetic information. Both can be reproduced indefinitely, providing the environment supports the use of that information. Nevertheless, alternative schemas/genes may be more appropriate for environmental conditions. If inappropriate, the structure of the information encapsulated in schemas or genes must change. Changes or variations are generated randomly and tested against the environment. If successful, a new schema or gene will be constructed and used in future. Thus, natural selection and the processing of information by human cognitive architecture can be characterized as identical ways of handling very complex information.

E. Generating Additional Matrices of Continua

This analysis suggests that the cognitive matrix of continua depicted in Fig. 1 is a specific example of a more general matrix from which examples such as that of Fig. 1 can be generated. If so, that more general matrix should be capable of generating not only the psychological example of Fig. 1, but also a specific example applicable to evolutionary biology. The ability to generate a general matrix from Fig. 1 and to generate, in turn, an example applicable to evolution would provide evidence for the argument that common information structures underlie human cognitive architecture and evolution by natural selection. Figure 2 depicts a general matrix of continua that can be used to generate specific matrices applicable to particular areas that may have the same underlying information structures. Figure 3 depicts the evolutionary example that can be derived from Fig. 2.

The first continuum of Fig. 1 deals with learning. On the left side of this continuum, we need to learn (or adapt) when we do not have knowledge

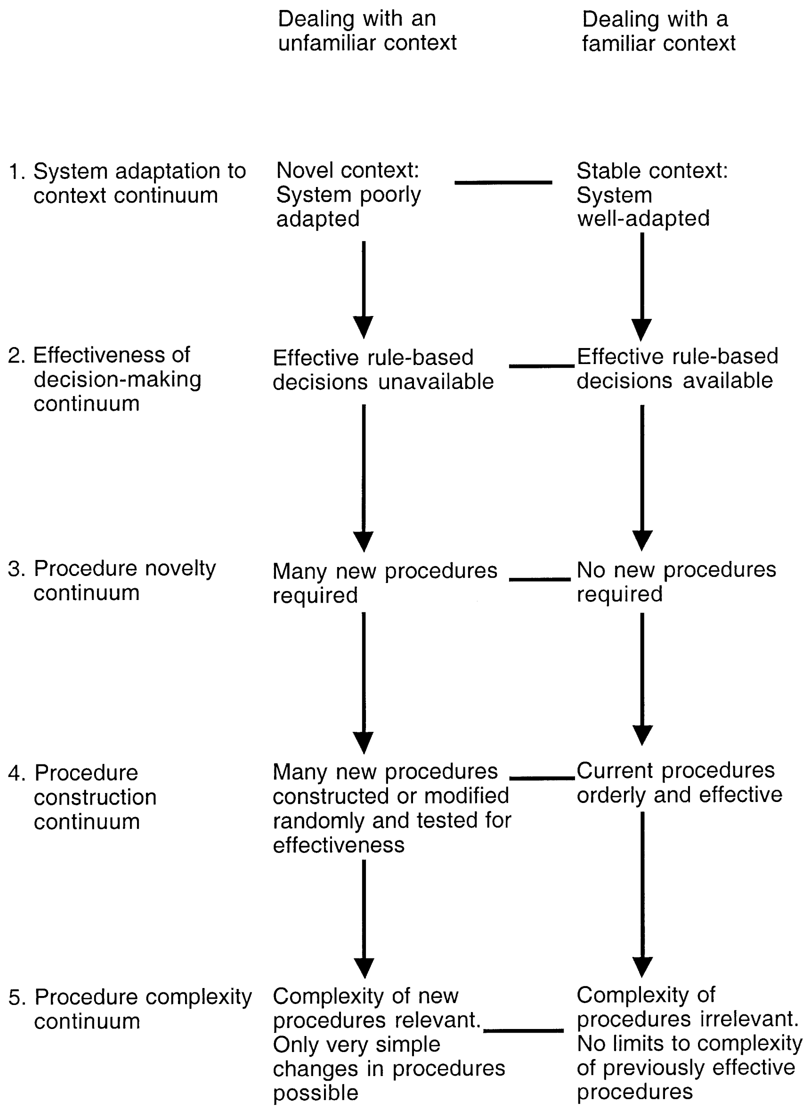

Fig. 2. A generalized matrix of continua.

needed to function in a particular environment. On the right side, essential knowledge has been acquired. In the more general terms of Fig. 2, on the left side, the first continuum deals with an information system that is operating in a novel context for which it is poorly adapted. It needs to adapt or

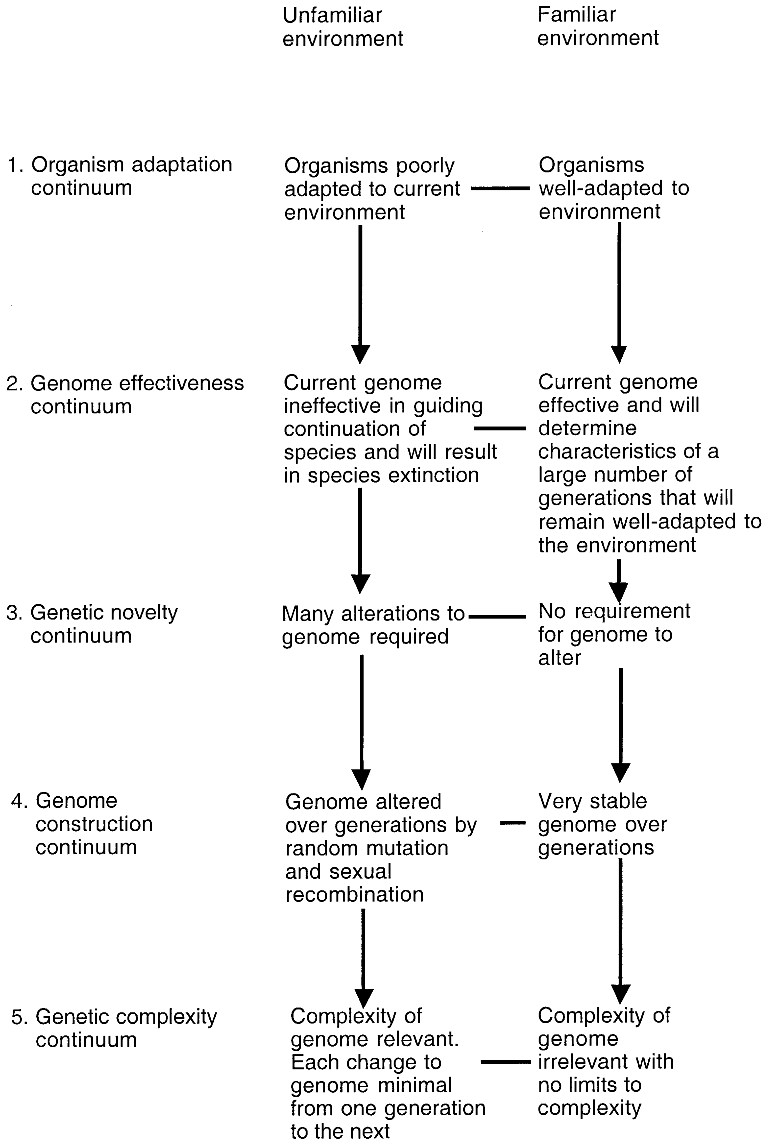

Fig. 3. A matrix of continua for evolution by natural selection.

"learn." On the right side, the system has already adapted or "learned" what is needed to operate in its environment. The first continuum of the specific evolution by the natural selection continuum of Fig. 3 varies from organisms that are poorly adapted to their current environment and so need to adapt to organisms that are well adapted to their environment.

The second continuum of each of the three figures is concerned with the extent to which performance is guided by established rules. In the case of Fig. 1, dealing with cognitive architecture, on the left side when faced with new material, there are no schemas to guide performance. On the right side, when dealing with familiar material, schemas determine actions. Thus, in the general terms of Fig. 2, on the left there are no available rules to govern the way the system should operate in its environment, whereas on the right there are well-established rules. This general continuum is the second continuum of Fig. 2. Translated into evolutionary terms, on the left we have a genetic endowment that will not permit a species to survive without change, whereas on the right we have a species with a genetic endowment that is well adapted to the current environment.

If a system is not adequately adapted to its environment, it needs to alter. The left side of the third continuum of Fig. 1 indicates that humans engage in problem solving when faced with such a situation. On the right, where material is well learned, adaptation or problem-solving search is unnecessary. The third continuum of Fig. 2 describes a general continuum in which at one extreme, many new procedures are required to permit the system to operate in the prevailing environment to a situation at the other extreme where no new procedures are required because the system is well adapted to the current circumstances. Similarly, in the genetic terms of the third continuum of Fig. 3, many alterations to the genome are required for survival on the left side of the continuum as opposed to no requirement for alterations to the genome on the right side.

If change is required, what are the mechanisms of change? For human cognitive architecture, the left side of the fourth continuum indicates that change occurs randomly. (Recall that while the generation of possible changes is random, assessment of the effectiveness of possible changes is not random.) On the right side of the continuum, change is not required because previously acquired schemas indicate what actions to take faced with a problem. In other words, we have a system that must generate new procedures randomly and test them for effectiveness at one extreme of the fourth continuum of Fig. 2 or is able to use currently established procedures at the other end of the continuum. In evolutionary terms, as depicted in the fourth continuum of Fig. 3, random mutation and sexual recombination are needed to generate changes to the genome and perhaps new species if a line is to survive. Alternatively, at the other end of the continuum, the current genome is satisfactory for survival without substantial alteration.

Finally, if elements are combined randomly, there must be mechanisms that ensure combinatorial explosions are kept in check. The limited working memory on the left side of the fifth continuum of Fig. 1 provides such a mechanism. In contrast, on the right side of the continuum, working memory limitations are not needed and do not occur because previous learning has ensured orderly and appropriate sets of elements irrespective of the size of those sets. In general terms of the fifth continuum of Fig. 2, if new procedures are being generated randomly, there must be mechanisms to limit their complexity. Changes must be relatively small and simple to reduce the number of possible changes and to reduce the probability that any change will result in a breakdown of the system. On the right side of the fifth continuum, procedures that are effective need have no limits to their complexity. In other words, while changes to the system must be small and incremental, there are no limits to the complexity of the resulting system. From the perspective of evolution by natural selection, while alterations to the genome from one generation to the next are minimal, as indicated on the left side of the fifth continuum of Fig. 3, that process, if permitted to continue for a sufficiently long period, can result in the immensely complex genome referred to on the right of the fifth continuum. There may be no limit to genetic complexity under such circumstances.

The isomorphism of Figs. 1, 2, and 3 provides evidence for the suggestion that human information processing recapitulates evolution by natural selection. They both share common information structures. It is understandable that the management of information by human cognitive architecture and evolution by natural selection should be similar. Evolution by natural selection is possibly the most efficient, natural system for transmitting, altering where necessary, and perpetuating information. It might be expected that human cognitive architecture, which must also manage information, would evolve to mimic the information processing procedures of evolution by natural selection because both systems are based on the general information processing procedures of Fig. 2.

III. Instructional Consequences

A. General Instructional Consequences

Instruction is only necessary toward the unlearned end of the learning continuum of the cognitive matrix of continua (Fig. 1), and one of its primary functions is to provide a partial substitute for the missing central executive at this end of the continuum. Consider again someone wishing to learn the road route from point A to point B. They can have someone explain the route, use a road map, or use a combination of prior knowledge with a problem-solving search to fill in the gaps. These activities function as a central executive in different ways and have different instructional consequences. Both an explanation and a map are two different forms of direct instruction, whereas problem solving provides an example of exploratory learning.

An explanation provides a strong substitute for a cognitive central executive. As one would expect from a central executive, it provides an overarching set of instructions for the critical processes that must be taken. Furthermore, the instructions can be followed with a minimum of additional learning, such as learning to use a map. If the explanations are adequate, all random processes are eliminated because the explanation, as a central executive, tells the learner precisely what needs to be done. Once a road route is learned, the learner moves to the right side of the matrix of continua, and the schemas acquired take over from the explanation and act as the central executive, rendering an explanation redundant.

A map, while it also acts as a substitute for a cognitive central executive, requires more intermediate learning than an explanation before it can be used. People need to learn to use a map before they can use it to learn a particular route. Thus, learning to use a map has its own set of learning continua, and until a person has acquired the map-reading schemas that allow movement to the right side of the matrix of continua for map reading, learning a route by using a map will be difficult or even impossible. Nevertheless, if map-reading skills have been acquired, a map, like explanations, can provide a powerful central executive substitute. Used properly the need to consider the consequences of random actions can be totally obviated and can continue to be avoided until the schema-based central executive on the right of the matrix of continua takes over the executive functions.

Problem solving provides the least effective substitute for a cognitive central executive. There is no choice but to propose actions randomly and then use the environment or prior knowledge to test the effectiveness of those actions as far as they can be tested. The learner is likely to move to the right of the matrix of continua very slowly, and so for much of the learning process, there is no effective central executive function. Only toward the end of the learning process, when schemas have been acquired, is an effective central executive available. Using this reasoning, problem solving may be considered as a last resort instructional technique when other more direct forms of instruction are unavailable.

The inadequate central executive function provided by problem solving has other ramifications. Combining elements randomly and testing the effectiveness of combinations against reality require substantial working memory resources (Sweller, 1988). The activity imposes a heavy working memory load just at the point where working memory resources are at their weakest because problem-solving search occurs at the new, yet-to-be-learned end of the learning continuum where working memory limitations are relevant. The heavy working memory load associated with problem solving can interfere with learning. Direct, fully guided instruction alternatives to problem solving circumvent both the lack of a central executive and the heavy cognitive load associated with search. On this analysis, direct guided instruction, rather than problem solving, should be used as a means of acquiring schemas. Substantial empirical evidence exists for this suggestion (see Sweller, 1999; Sweller, van Merrienboer, & Paas, 1998; Tuovinen & Sweller, 1999).

The contrast between direct guided instruction and exploration applies to all material that needs to be learned, including material covered in educational institutions. Learning to solve classes of mathematical problems, write essays in history, run scientific experiments, or learning to read and write must all be affected without an adequate cognitive central executive provided by schemas. Showing students how to solve mathematical problems, write particular types of essays, run experiments, or providing direct instruction in how to read and write can all provide an effective central executive substitute and reduce the cognitive load associated with problem solving, although care must be taken to ensure that the instruction itself does not impose a heavy working memory load (e.g., Sweller, Chandler, Tierney, & Cooper, 1990; Sweller, Mawer, & Ward, 1983). In all cases, direct guided instruction can provide a temporary replacement for schemas until they are acquired.

Indirect instruction provided by various discovery/exploratory techniques offers a less effective central executive substitute with an inevitably high random component. Direct guided instruction is effective because it reduces the number of random element combinations that must be tested. It is likely to be essential for very high element interactivity material for which the number of random combinations that must be tested will be unacceptably high. The knowledge that lies behind such material could only be derived by scholars engaged in the very lengthy, working memory-taxing activities inevitably required when dealing with a multitude of interacting elements that are not appropriately organized by a central executive. Such problem-solving activity is unavoidable when neither schemas nor direct instruction are available to act as a central executive that indicates appropriate relations between elements. Humans learn through problem solving not because it is effective (empirical evidence indicates unambiguously that it is not effective as a learning device, see Sweller, 1999; Sweller et al., 1998) but because they are forced to by the environment and the lack of a central executive. Direct guided instruction acts as a substitute for a central executive and should always be used if available.

B. Creativity

Creativity has always been a difficult concept to deal with or even to define. Nevertheless, most definitions of creativity incorporate the generation of new ideas and, under such definitions, it is easy to assume that the general instructional consequences discussed in the previous section leave no room for human creativity or may even stifle creativity. In fact, the common information processing structures of human cognitive architecture and evolution by natural selection can provide a solution to the problem of human creativity.

Evolution by natural selection has created innumerable functions, procedures, and outcomes that vastly exceed the capability of human cognition. We are not only unable to create what evolution by natural selection has created, to this point we are unable to even understand many of the products of evolution, with massive scientific enterprises being devoted to precisely this cause. Given the much shorter time frame in which human cognitive activity operates, it is not surprising that our creative endeavors are unable to match those of evolution by natural selection. Nevertheless, humans are and have been creative and that creativity can be explained by the current theoretical framework. Based on the perspective of this chapter, human creativity and the creativity exhibited by evolution by natural selection are generated by the same mechanisms. Those mechanisms are reflected on the left side of the matrices of continua. A knowledge base in long-term memory or as part of a genetic code may become inadequate and is altered by random processes; the knowledge base requires procedures for testing the effectiveness of alterations and only incorporating those that are effective; and the knowledge base must have mechanisms to protect it from large random alterations that may destroy it. Using these mechanisms, both evolution by natural selection and human cognition have been able to create new and effective structures.

It needs to be noted that on this analysis, random processes provide the initial impetus for human creativity just as random mutation is critical for the creativity of evolution by natural selection. There is no central executive determining what is creative (left-hand side of the second continuum of Figs. 1 and 3). Nevertheless, despite the initiating random processes, creativity is critically determined by the current knowledge base, as it is from that base that new creative actions are taken, just as it is the information encapsulated in a genome from which random mutations can determine new biological procedures and functions (fourth continuum of Figs. 1 and 3).

Langley, Simon, Bradshaw, and Zytkow (1987) also suggested that creativity depends on an appropriate knowledge base associated with conventional problem-solving search mechanisms. Some evidence for the validity of their proposal comes from a production system that they constructed that rediscovered some of the early laws of physics. That system only had the knowledge base required to generate particular laws and so has not been able to discover new scientific laws. If the theoretical suggestions made in the current chapter are valid, no computational system is likely to discover, as opposed to rediscover, new scientific laws unless it incorporates a massive knowledge base with the mechanisms for small random alterations of that base over long periods of time along with procedures for testing the effectiveness of those alterations. Such a system is currently not available.

Suggested procedures for "teaching" creativity arise periodically in both psychology and education. None of these attempts has been able to obtain widespread, empirical support. The current proposals imply that teaching creativity is likely to be difficult or impossible but that humans may no more need to be taught how to "explore," "investigate," "discover," or "create" than does evolution by natural selection. Only a knowledge base can be taught and learned and that knowledge base will determine what can and cannot be created.

It is, of course, possible that life on earth includes multiple mechanisms that have creativity as one of their end results and that the creativity exhibited by evolution by natural selection and by humans uses different mechanisms. Nevertheless, the thesis outlined in this chapter suggests a single rather than multiple mechanism.

C. Specific Instructional Design Principles and Effects

There are a range of specific instructional design principles and effects that flow from the considerations outlined in this chapter. Cognitive load theory, an instructional theory based on the combination of information structures and cognitive architecture described earlier, has been used to generate those instructional effects.

1. *The Goal-Free Effect*

This effect occurs when learners presented a conventional, goal-specific problem in which the goal might be "calculate the value of angle ABC" in the case of a geometry problem or "calculate the final velocity of the vehicle" in the case of a kinematics problem learn less than learners presented a nonspecific or goal-free problem. Examples of nonspecific goal problems are "calculate the value of as many angles as you can" or "calculate the value of as many variables as you can." The goal-free effect was obtained by Sweller and Levine (1982) and has been demonstrated on many occasions (Ayres, 1993; Sweller, & Cooper, 1985; Burns & Vollmeyer, 2002; Geddes & Stevenson, 1997; Miller, Lehman, & Koedinger, 1999; Owen & Sweller, 1985; Paas, Camp, & Rikers, 2001; Sweller, 1988; Sweller et al., 1983; Tarmizi & Sweller, 1988; Vollmeyer, Burns, & Holyoak, 1996). It can be explained using the cognitive matrix of continua of Fig.1.

Assume a novice problem solver solving conventional problems by means-ends analysis. As a novice, he or she will be on the left side of the matrix of continua. To make moves, differences between the current state and the goal state will need to be established, a potential move will need to be chosen randomly (assuming prior knowledge concerning the effects of particular moves is unavailable), and each potential move will need to be assessed to establish whether it reduces differences between the current problem state and the goal state. Because working memory limitations are relevant on the left side of the matrix of continua, this complex procedure may leave few or no resources available to attend to schema acquisition. When acquiring a schema, learners must engage in the quite different activity of learning to classify problems and problem states according to their moves. As a consequence, learning may be inhibited.

In contrast, assume a problem solver who is presented goal-free problems. The only activity that needs to be engaged in is to choose any potential moves randomly and determine whether they can be made. Working memory load is minimal. Furthermore, learning which moves can be made given a particular problem state is central to schema acquisition. Sweller (1988) suggested that this interpretation explains the goal-free effect.

Presenting learners with goal-free problems may appear unusual if the aim is to present learners with direct, fully guided instruction. Goal-free problems reduce the guidance provided by a specific goal. For this reason, the procedure is effective, but only if all moves made under goal-free conditions are useful in the sense that they need to be learned and practiced. Not all problems have this characteristic. Some problems have a large or even infinite number of moves that could be made with most moves serving no function. For example, asking learners to make as many manipulations as possible of the equation $(a + b)/c = d$ can result in an infinite number of manipulations, as one can legitimately add an infinite number of constants to each side. Goal-free problems should not be used with such material and so an alternative is required.

2. *The Worked Example Effect*

The use of worked examples can overcome the problem of goal-free problems only being useful for a limited class of problems. There are probably no classes of problems for which worked examples are not potentially effective.

The worked example effect occurs when learners who are presented with a large number of worked examples to study learn more than learners presented an equivalent number of problems to solve. The effect has been studied extensively (Carroll, 1994; Cooper & Sweller, 1987; Miller et al., 1999; Paas, 1992; Paas & van Merrienboer, 1994; Pillay, 1994; Quilici & Mayer, 1996; Sweller & Cooper, 1985; Trafton & Reiser, 1993).

Worked examples provide problem-solving guidance that can act as a substitute for schemas that are unavailable to novices. They are the ultimate form of direct instruction. Rather than engaging in the means-ends problem-solving search process described earlier, learners can be guided by a worked example acting as a substitude schema-based central executive. The lack of such a central executive necessitates problem-solving search, with its inevitable random components and working memory load found on the left side of the matrix of continua. While psychologically the learner is on the left side of the matrix of continua, a worked example allows him or her to perform as though they are on the right side of the matrix. A good example acts as a substitute for a schema-based central executive, eliminates the problem-solving search with its random base, and reduces difficulties imposed by a limited working memory because all necessary information is incorporated within the example (see later sections on split-attention, modality, and redundancy effects). As a consequence, learning can be facilitated by an emphasis on worked examples resulting in the worked example effect.

3. *The Problem Completion Effect*

Most demonstrations of the worked example effect involve presenting worked examples paired with very similar problems. Learners are presented a worked example and are then immediately presented a very similar problem to solve. This procedure ensures that learners are motivated to study the worked example in order to ensure that they can solve the following problem. The extent to which they can solve the following problem also provides them with some feedback concerning their ability to solve problems of that type.

Completion problems were invented as an alternative to this procedure. Rather than presenting learners with full worked examples followed by similar problems, they are presented with partial worked examples that require completion. The partial worked example provides sufficient guidance to reduce the problem-solving search and cognitive load, whereas problem completion ensures that learners are motivated and receive feedback concerning their knowledge of relevant problem types. Paas (1992), Paas and van Merrienboer (1994), van Merrienboer (1990), van Merrienboer and de Croock (1992), van Merrienboer and Krammer (1987), and van Merrienboer, Schuurman, de Croock, and Paas (2002) provided evidence that completion problems have a positive effect similar to that of worked examples when compared to full problems. It is reasonable to assume that the theoretical reasons for the problem completion effect are identical to those used to explain the worked example effect.

4. *The Split-Attention Effect*

Not all worked examples are effective. A worked example that is structured in a manner that ignores working memory limitations may be no more or even less effective than solving the equivalent problem. Some worked examples in some areas are conventionally structured in a manner that requires learners to split their attention between multiple sources of mutually referring information before mentally integrating those sources of information. A conventional geometry worked example consisting of a diagram and statements provides an instance. The diagram in isolation provides no instruction. The associated statements, such as *angle ABC* = *angle XYZ*, are unintelligible without a diagram. Meaning can only be derived from the worked example by mentally integrating the diagram and the statements. Mental integration requires working memory resources because learners must search for referents. When a geometry statement refers to *angle ABC*, learners must search the diagram for *angle ABC* in order to understand the statement. In effect, the learner is not only on the left side of the matrix of continua for geometry, but is on the left side of the matrix for the particular example being studied. An act of problem solving must be engaged in simply to locate appropriate referents. Locating referents requires working memory resources that are unavailable for learning geometry.

Because we do not normally have schemas for the labeling of particular geometry diagrams, providing guidance in locating referents can be just as beneficial as guidance in the more general terms discussed previously. Such guidance can be provided by physically integrating diagrams and statements. Rather than placing the statement *angle ABC* = *Angle XYZ* below or next to the diagram as normally occurs, the relevant statements can be incorporated within the diagram so that a search for referents is eliminated. If conventionally structured worked examples are compared with physically integrated examples, results normally demonstrate an advantage for the integrated versions, resulting in the split-attention effect. Various versions of the effect have been demonstrated using a wide variety of materials under a wide variety of conditions. Furthermore, as might be expected, it is not restricted to worked examples but applies to any instructional material (Bobis et al., 1993; Cerpa, Chandler, & Sweller, 1996; Chandler & Sweller, 1992, 1996; Mayer & Anderson, 1991,1992; Mwangi & Sweller, 1998; Sweller et al., 1990; Tarmizi & Sweller, 1988; Ward & Sweller, 1990).

5. *The Modality Effect*

While physical integration of multiple sources of information can be highly effective, there is an alternative that is equally effective and, under some circumstances, may be preferable. The split-attention effect relies on visual modality with visual search being reduced by the use of physical integration. Visual search means that the visual channel only (the visuospatial sketch pad of Baddeley, 1992; Baddeley & Hitch, 1974) is being used and overloaded under split-attention conditions. Considerable evidence, shows that effective working memory can be increased by using dual rather than a single modality (e.g., Penney, 1989). While the visual and auditory processors of working memory are not fully separate in the sense that one does not obtain a simple additive increase in processing capacity by presenting some material visually and some in auditory mode, there is considerable empirical evidence of a measurable increase in working memory capacity when using both modalities (Allport, Antonis, & Reynolds, 1972; Brooks, 1967; Frick, 1984; Levin & Divine-Hawkins, 1974). From a theoretical perspective, capacity should increase to the extent that visual and auditory processors can function autonomously without sharing other cognitive structures that limit capacity. Some empirical evidence of an increase in working memory capacity when using both modalities also provides evidence for partial autonomy of the auditory and visual channels.

The possibility of increasing working memory capacity using dual rather than a single modality should have instructional consequences. For example, under split-attention conditions, rather than presenting a diagram and written text that should be integrated physically, it may be possible to present a diagram and spoken text. Because the diagram uses visual modality while speech uses auditory modality, the total available working memory capacity should be increased, resulting in enhanced learning.

The instructional modality effect occurs when learners, faced with two sources of information that refer to each other and are unintelligible in isolation, learn more when presented with one source in visual mode and the other in auditory mode rather than both in visual mode. This effect has been demonstrated on a number of occasions (Jeung, Chandler, & Sweller, 1997; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Mousavi, Low, & Sweller, 1995; Tindall-Ford et al., 1997).

6. *The Redundancy Effect*

Both split-attention and modality effects occur under very specific conditions. They are only obtainable when multiple sources of information refer to each other and are unintelligible in isolation. For example, a diagram and text will not yield either split-attention or modality effects if the diagram is fully intelligible and fully provides the information needed, with the text merely recapitulating the information contained in the diagram in a different form. Under such circumstances, the text is redundant. The redundancy effect occurs when additional information, rather than having a positive or neutral effect, interferes with learning. For example, instead of integrating a diagram with redundant text or presenting the text in auditory form, learning is enhanced by eliminating the text.

There are many different forms of redundancy. The previous diagram/text redundancy occurs when a self-explanatory diagram has additional text redescribing the diagram (Chandler & Sweller, 1991). Mental activity/physical activity redundancy occurs when, for example, learning how to use a computer application by reading a text has the added physical activity of using the computer (Cerpa et al., 1996; Chandler & Sweller, 1996; Sweller & Chandler, 1994). Either reading the text in a manual or, surprisingly, physically using a computer can be redundant and interfere with learning. Summary/detailed exposition redundancy occurs when a summary alone results in enhanced learning compared to a full exposition (Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Reder & Anderson, 1980, 1982) Finally, auditory/visual redundancy occurs when the same material, presented simultaneously in written and spoken form, results in a learning decrement compared to the material presented in written or auditory form alone (Craig, Gholson, & Driscoll, 2002; Kalyuga, Chandler, & Sweller, 1999, 2000; Mayer, Heiser, & Lonn, 2001).

The redundancy effect is one of the more surprising cognitive load effects, with many people finding it quite counterintuitive. Most of us feel that even if additional explanatory material is not beneficial, at the very least it should be neutral. In fact, the addition of redundant information can have strong, negative consequences. The effect can be understood in cognitive load theory terms. If one form of instruction is intelligible and adequate, providing the same information in a different form will impose an extraneous cognitive load. Working memory resources will need to be used to process the additional material and possibly relate it to the initial information. It is likely to be only after the learner has processed the additional information that he or she will be aware that the activity was unnecessary. At that point, the damage may have been done.

7. *The Element Interactivity Effect*

Split-attention, modality, and redundancy effects all occur as a consequence of instructional procedures designed to reduce working memory load. It might be expected that the instructional procedures would only be effective where the material being learned imposed an intrinsically high cognitive load. If material does not impose a high cognitive load, the additional load due to inadequate instructional procedures may not matter a great deal because working memory capacity may not be exceeded. An extraneous cognitive load due to inadequate instructional procedures may be irrelevant if the intrinsic cognitive load imposed by the structure of the information is low. Because low element interactivity material has a low intrinsic cognitive load, we can predict that cognitive load effects may disappear when learning such material. The effects may only be obtainable using high element interactivity material. Chandler and Sweller (1996) and Sweller and Chandler (1994) demonstrated that split-attention and redundancy effects could be demonstrated readily using high element interactivity material but disappeared when low element interactivity material was used. Tindall-Ford et al. (1997) similarly found that the modality effect could only be obtained using high element interactivity material. Marcus et al. (1996) found that diagrams for which we have schemas facilitated understanding when compared to text but only under conditions of high element interactivity.

The finding that cognitive load effects can only be obtained using high element interactivity material demonstrates the element interactivity effect. It consists of an interaction between the split-attention, redundancy, and modality effects and the complexity (as measured by element interactivity) of the material being learned. While it has not been tested using other cognitive load effects, there is every reason to suppose that it could be obtained with all other effects based on a limited working memory.

It has been suggested in this chapter that the particular interaction between a limited working memory and a very large long-term memory had to evolve in order to handle high element interactivity material. High element interactivity material must be imbedded in schemas before it can be handled by a limited working memory. The element interactivity effect indicates that when instruction deals with high element interactivity material, then the characteristics of human cognitive architecture, such as a limited working memory, become critical.

8. *The Isolated Interacting Elements Effect*

Consider a learner faced with new material. That learner is on the left side of the cognitive matrix of continua. Consider further that element interactivity of the information that must be assimilated is sufficiently high to substantially exceed working memory capacity. Understanding cannot occur because understanding requires all interacting elements to be processed simultaneously in working memory. All the interacting elements cannot be processed simultaneously in working memory until schemas have been formed, but schemas will not be formed until the learner has moved toward the right of the matrix of continua. Because the learner cannot possibly understand the material until those schemas have been formed, understanding and learning may appear impossible at first sight. When the material is presented with all of its interacting elements, as it needs to be if understanding is to occur, it cannot be processed in working memory because it vastly exceeds working memory capacity. How does learning occur under such circumstances?

One possibility (perhaps the only possibility) is that initially the elements must be learned as though they are isolated, noninteracting elements. Once sufficiently sophisticated schemas have been constructed, understanding will occur because the interacting elements can now be processed in working memory. On this analysis, learning must precede understanding.

If this analysis is valid, it is reasonable to hypothesize that learning might be facilitated by initially presenting very complex information to students in isolated elements form without emphasizing or even indicating the interactions between elements. Understanding of such instruction will be limited, but once it has been learned, presentation of the full information may permit understanding to occur. In contrast, presentation of the complete information that potentially could be understood during initial instruction may result in very little learning or understanding. Pollock et al. (2002) obtained precisely this effect. Learners presented isolated elements to learn followed by the full set of interacting elements learned more than learners presented the full set of interacting elements twice, demonstrating the isolated interacting elements effect.

9. *The Imagination Effect*

Assume a novice on the left of the cognitive matrix of continua has acquired some schemas and is beginning to move toward the right of the continua. To attain relatively high levels of expertise, further learning will need to include automation of the previously acquired schemas that normally includes continuing to study the material until desired levels of performance have been attained. An alternative is to attempt to imagine the procedures that have been learned. Imagining requires the learner to mentally "run through" the procedures in working memory. For high element interactivity material, processing information in working memory is impossible until schemas have been acquired. Once they have been acquired and the learner has moved toward the right of the matrix of continua, imagination techniques should be feasible and practice through imagination should assist in automation. Continuing to study the material should be unnecessary because studying high element interactivity material is designed to provide the guidance necessary to reduce search while acquiring schemas, as occurs on the left side of the matrix of continua. If schemas have already been acquired, there is no longer any need to provide instructional guidance to reduce search because, on the right of the matrix of continua, the central executive function of schemas is now able to operate. Using those schemas to imagine the procedures learned should facilitate further learning through automation in a manner that studying the instructions may not.

Cooper, Tindall-Ford, Chandler, and Sweller (2001) tested this hypothesis and found that learners given instructions to "imagine" a set of procedures that needed to be learned performed better than learners given conventional "study" instructions. This imagination effect was only obtained using learners with sufficient knowledge to be able to process all of the required information in working memory. For complete novices who were unable to process the high element interactivity material in working memory, a reverse imagination or "study" effect was obtained with "study" instructions proving superior to "imagination" instructions. In other words, the effect obtained depended on the levels of expertise of the learners. Higher levels of expertise could reverse the effect obtained. The ideal form of instruction depended on the expertise of the learners. This reversal effect with expertise, as it happens, is general. As described in the next section, most, perhaps all, of the cognitive load effects described earlier depend on the use of novices.

10. The Expertise Reversal Effect

With the exception of the imagination effect, all of the previously described effects were intended to provide new instructional procedures for novices who were on the far left of the cognitive matrix of continua. Learners, of course, continue to learn and may require instructional procedures after they have moved beyond the left of the matrix of continua. It turns out that frequently, once learners have acquired some knowledge, many of the effects described previously reverse. With increased experience, conventional instructional procedures, such as practice at solving problems, are better than cognitive load procedures, such as studying worked examples. The imagination effect differs from the other effects discussed in that the imagination technique is intended for more knowledgeable learners rather than complete novices and so reverses when the imagination technique is presented to novices rather than the more experienced learners. In all other cases, the effects shown using novices are reversed when using more experienced learners. The reversal is due to the redundancy effect and is called the expertise reversal effect. It is due to an interaction between simpler cognitive load effects and levels of expertise and can be contrasted with the element interactivity effect, discussed earlier, which consists of an interaction between simpler cognitive load effects and task complexity.

Using diagrams and text, Kalyuga, Chandler, and Sweller (1998) obtained the normal split-attention effect with integrated diagrams and text proving superior to a split-attention format. A group presented the diagrams alone performed poorly because the text was essential in helping understand the diagram, a necessary condition for the split-attention effect. The learners used were novices on the left side of the cognitive matrix of continua. Over several months training in the relevant, engineering area, they moved toward the right of the matrix of continua. The necessary guidance provided by the text became less and less essential as schemas were acquired to take over from the text. The superiority of the integrated format decreased with increased expertise. A point was reached where there was no difference between groups. Eventually, with additional training, the text became redundant. Learners could understand and learn from a diagram alone. Having to process unnecessary text increased the cognitive load. The presence of redundant text, especially in integrated form where it is difficult to ignore, interfered with rather than facilitated learning. A redundancy effect was obtained with the diagram-alone condition providing the best learning environment.

Yeung, Jin, and Sweller (1998) obtained a similar effect using textual materials. Learners with low levels of language competence were assisted by explanatory notes integrated into the primary text. Integrated notes retarded learning for learners with higher levels of language competence because the notes were redundant but were difficult to ignore when integrated into the primary text.

Other cognitive load effects also disappear and then reverse with increased expertise. A modality effect obtained with novices disappeared and then reversed (Kalyuga, Chandler, & Sweller, 2000) as expertise increased. Novices required textual material to assist them understand visually presented material; that textual material was best presented in spoken rather than written form, demonstrating the modality effect. As expertise increased, that modality effect disappeared and eventually, presenting the visual material alone was superior to an audiovisual presentation or, indeed, any presentation that included the text. Guidance provided by textual material, essential to students on the left of the cognitive matrix of continua, was provided by the schemas now available after students had moved to the right side of the matrix.

Similarly, Kalyuga, Chandler, Tuovinen, and Sweller (2001) found that the worked example effect reversed with increased expertise. Novices require worked examples to provide them with guidance. Schemas, once they have been acquired, provide guidance, and worked examples become redundant. Kalyuga, Chandler, and Sweller (2001) and Tuovinen and Sweller (1999), using novices, found that direct instruction is superior to discovery learning. That difference disappeared if learners with more experience in the domain were used.

These results can be used to explain other findings. McNamara, Kintsch, Songer, and Kintsch (1996) found that when learners were presented a textual passage to read and assimilate, those who were relatively expert in the area learned more from reduced passages that had segments omitted than the full passage. Learners with less experience in the area learned most using the full passage. On the present interpretation, novices required the full passage to allow understanding and so the full passage condition was superior. With increased experience, the added material was redundant and merely served to obscure critical points. Working memory resources were required to extract those critical points from the surrounding, redundant material, reducing learning and resulting in the superiority of the reduced passage.

11. The Guidance Fading Effect

From an instructional perspective, the expertise reversal effect suggests that as learners move from the left of the cognitive matrix of continua to the right, schemas increasingly provide guidance and so the guidance provided by instruction should be faded out. Unnecessary guidance has negative, not simply neutral effects. Renkl and associates (Renkl, 1997; Renkl, Atkinson, & Maier, 2000) obtained precisely this result using combinations of worked examples, completion problems, and full problems. Using novices, they found that guidance provided by worked examples was the best form of instruction. With increasing expertise, it was desirable for those worked examples to be replaced with completion problems and, ultimately, with full unguided problems.

It was indicated earlier that for novices, instruction should replace the missing central executive but that with increased levels of expertise, schemas play the role of a central executive. A guidance fading technique accords closely with this suggestion. Initially, with no central executive available, worked examples indicate relations between elements of information. As rudimentary schemas begin to form, they can take over some of the central executive function from worked examples and so complete worked examples are no longer necessary. Completion problems can be used as a substitute for worked examples. Once full schemas have been constructed, they can act as a central executive and so full problem solving with no other guidance can be instituted. Additional learning through schema automation should occur during problem-solving practice.

Renkl, Atkinson, Maier, and Staley (2002) found guidance fading as levels of expertise increase to be demonstrably superior to using a single instructional procedure. They compared the presentation of conventional worked examples with guidance fading. The worked example procedure incorporated the presentation of several pairs consisting of a worked example followed by a very similar problem to solve. This pairing of a worked example followed by a problem was used throughout the learning period, irrespective of changing levels of expertise. Results indicated that the guidance fading procedure was superior. The superiority of fading over a single design procedure (e.g., worked examples alone or problems alone) as expertise increases constitutes the guidance fading effect.

The guidance fading effect, along with the expertise reversal effect, indicates that individual differences, specifically differences in levels of expertise, are a critical consideration when choosing an instructional design. A design that is ideal for a person located toward the left of the cognitive matrix of continua may be quite inappropriate for someone further to the right of the matrix. Ignoring levels of expertise can result in the use of quite inappropriate instructional procedures.

The instructional designs described in this section differ from most instructional designs in that they are very closely tied to our knowledge of information structures and human cognitive architecture. Indeed, they were generated directly from that knowledge. It can be argued that they provide a degree of validity to the cognitive theories discussed. In any scientific area, it is difficult or impossible to generate applications from substantially faulty theories.

IV. Conclusions

Human cognitive architecture has evolved to permit humans to engage in activities that range from prosaic to awe inspiring. There are logical structures that determine the way in which cognitive architecture deals with information. Those logical structures, along with the structure of information itself, must have determined the course of the evolution of human cognitive architecture. The basic information structures that underlie human cognitive architecture consist of a very large information store with limitations to ensure that any changes to that store do not destroy its functionality. The end result is an architecture designed to store immeasurable amounts of information in a long-term memory but a very limited ability to deal with novel information in working memory. Information held in long-term memory guides most of our activities. Novel information in working memory can feed information into long-term memory and so alter long-term memory, but the logic of the governing information systems ensures that the alterations are relatively small to circumvent the unavoidable random components.

As might be expected, this system logic is universal. It not only applies to the manner in which human cognitive architecture has evolved, it applies to the manner in which information is handled by the processes of evolution themselves. Evolution by natural selection can be characterized as an effective and efficient system for managing and adapting very complex, natural information to changing circumstances. Human cognitive architecture must also manage complex information. Accordingly, it would not be surprising if human cognitive architecture evolved to handle information in the same way as evolution by natural selection. Similarities in the way that the two systems function suggest that human cognitive architecture, by the processes of evolution by natural selection, has itself evolved to duplicate the manner in which evolution by natural selection deals with information.

The logic of these systems places both restrictions on and generates opportunities for the manner in which information is presented and the activities in which learners should engage. Our cognitive architecture is structured with schemas providing an executive function guiding our mental activities. Instruction is required when those schemas are unavailable and must be acquired. Ideally, that instruction should provide an executive function that mimics the missing schemas as closely as possible in order to avoid random activities and reduce working memory load. Many instructional procedures that meet these requirements have now been devised. The successful generation of instructional procedures from theoretical principles provides a degree of validity for those principles.

While the logic of the information systems discussed in this chapter places immense barriers to their alteration, their adaptability to new circumstances, even if slow and frequently ineffective, is their crowning glory. Evolution may occur over eons but its whole point is change and adaptability, resulting in the creation of new functions, processes, and entities. Similarly, learning is the adaptive engine of human cognitive architecture. It may take many years, especially if creativity is required because instruction from and imitation of other humans is unavailable, but it is the foundation function of human cognitive architecture. Only through learning does the ability to efficiently process high element interactivity material become possible, and processing high element interactivity material is characteristic of humans. Prior to learning, such material can be dealt with but only in an unguided, partially random manner with all complex interactions ignored. Furthermore, there is an inevitability about this limitation. There can be no mechanism to coordinate the very large number of possible combinations that can occur when dealing with even a relatively small number of elements that have not been learned. Because knowledge acquired through learning provides the only coordinating function, it is essential that our cognitive architecture evolved to ensure that only a limited amount of uncoordinated information is considered at any given time prior to learning. This limitation creates an immediate tension when dealing with high element interactivity information that cannot be limited or reduced in size without compromising understanding. Because high element interactivity material must be coordinated, a mechanism for coordinating such information had to evolve if it was to be processed. Schematic knowledge acquired through learning is that mechanism. There are very wide or perhaps no limits to the amount of previously learned information that humans can process.

On this analysis, long-term memory is the source of human intellectual skill because long-term memory holds learned material. It may be this structure that took millions of years to evolve, and at least on earth, is unique to humans in terms of size. Our huge knowledge base is shared neither by other living creatures nor, to this point, by artificial devices created by humans. It may only be shared by the mechanisms that permit life itself to reproduce and evolve. Other cognitive structures, including ones not considered in this chapter, such as sensory systems, are both ubiquitous and frequently superior to their human equivalent. In contrast, our immense long-term memory, with its close connections to learning, has no cognitive equivalent on earth. That structure is quintessentially human.

Acknowledgments

The author thanks Paul Ayres, Brett Hayes, Paul Ginns, Slava Kalyuga, and Nadine Marcus for providing comments on earlier versions of this chapter. This work was supported by grants from the Australian Research Council.

References

Allport, D., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*, 225–235.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Erlbaum.

Aparicio, S. (2000). How to count ··· human genes. *Nature Genetics, 25*, 129–130.

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

Ayres, P. (1993). Why goal-free problems can facilitate learning. *Contemporary Educational Psychology, 18*, 376–381.

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

Baddeley, A., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology*. London: Cambridge University Press.

Ben-Zeev, T., & Star, J. (2001). Spurious correlations in mathematical thinking. *Cognition and Instruction, 19*, 253–275.

Bobis, J., Sweller, J., & Cooper, M. (1993). Cognitive load effects in a primary school geometry task. *Learning and Instruction, 3*, 1–21.

Bobis, J., Sweller, J., & Cooper, M. (1994). Demands imposed on primary-school students by geometric models. *Contemporary Educational Psychology, 19*, 108–117.

Britton, B., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*, 329–345.

Brooks, L. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology, 19*, 289–235.

Burns, B., & Vollmeyer, R. (2002). Goal Specificity effects on hypothesis testing in problem solving. *Quarterly Journal of Experimental Psychology, 55A*, 241–261.

Carroll, W. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*, 360–367.

Cerpa, N., Chandler, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research, 15*, 345–367.

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

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233–246.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 151–170.

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 ntelligence* (pp. 7–75). Hillsdale, NJ: Erlbaum.

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

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

Craig, S., Gholson, B., & Driscoll, D. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*, 428–434.

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

Dirlam, D. (1972). Most efficient chunk sizes. *Cognitive Psychology, 3*, 355–359.

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

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71–99.

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

Fingerman, P., & Levine, M. (1974). Non-learning: The completeness of the blindness. *Journal of Experimental Psychology, 102*, 720–721.

Frick, R. (1984). Using both an auditory and a visual short-term store to increase digit span. *Memory and Cognition, 12*, 507–514.

Geddes, B., & Stevenson, R. (1997). Explicit learning of a dynamic system with a non-salient pattern. *Quarterly Journal of Experimental Psychology, 50A*, 742–765.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.

Halford, G., Wilson, W., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 803–863.

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*, 860–921.

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.

Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*, 329–343.

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

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351–371.

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. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*, 5–23.

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.

Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition, 56*, 263–269.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126*, 278–287.

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

Langley, P., Simon, H., Bradshaw, G., & Zytkow, J. (1987). *Scientific discovery: Computational explorations of the creative process*. Cambridge, MA: MIT Press.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). Models of competence in solving physics problems. *Cognitive Science, 4*, 317–348.

Levin, J., & Divine-Hawkins, P. (1974). Visual imagery as a prose-learning process. *Journal of Reading Behavior, 6*, 23–30.

Levine, M. (1971). Hypothesis theory and non-learning despite ideal S-R reinforcement contingencies. *Psychological Review, 78*, 130–140.

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

MacGregor, J. (1987). Short-term memory capacity: Limitation or optimization? *Psychological Review, 94*, 107–108.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*, 49–63.

Mayer, R., & Anderson, R. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484–490.

Mayer, R., & Anderson, R. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*, 444–452.

Mayer, R., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*, 64–73.

Mayer, R., Heiser, J., & Lonn, S. (2001). Cognitive contraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187–198.

Mayer, R., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual-processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320.

McNamara, D., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1–43.

Miller, C., Lehman, J., & Koedinger, K. (1999). Goals and learning in microworlds. *Cognitive Science, 23*, 305–336.

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

Moreno, R., & Mayer, R. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.

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.

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction, 16*, 173–199.

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

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11–28.

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology, 77*, 272–284.

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., Camp, G., & Rikers (2001). Instructional compensation for age-related cognitive declines: Effects of goal specificity in maze learning. *Journal of Educational Psychology, 93*, 181–186.

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 and Cognition, 17*, 398–422.

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

Piaget, J. (1928). *Judgement and reasoning in the child.* New York: Harcourt.

Pillay, H. (1994). Cognitive load and mental rotation: Structuring orthographic projection for learning and problem solving. *Instructional Science, 22*, 91–113.

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

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*, 144–161.

Reder, L., & Anderson, J. R. (1980). A comparison of texts and their summaries: Memorial consequences. *Journal of Verbal Learning and Verbal Behavior, 19*, 121–134.

Reder, L., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for main points of a text. *Memory and Cognition, 10*, 97–102.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*, 1–29.

Renkl, A., Atkinson, R. K., & Maier, U. H. (2000). From studying examples to solving problems: Fading worked-out solution steps helps learning. In L. Gleitman and A. K. Joshi (Eds.), *Proceeding of the 22nd Annual Conference of the Cognitive Science Society* (pp. 393–398). Mahwah, NJ: Erlbaum.

Renkl, A., Atkinson, R., Maier, U., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education, 70*, 293–315.

Ross, B., & Kilbane, M. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 427–440.

Schneider, W., & Shiffrin, R. (1977). Controlled and automatic human information processing. I. Detection, search and attention. *Psychological Review, 84*, 1–66.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.

Sweller, J. (1980a). Transfer effects in a problem solving context. *Quarterly Journal of Experimental Psychology, 32*, 233–239.

Sweller, J. (1980b). Hypothesis salience, task difficulty, and sequential effects on problem solving. *American Journal of Psychology, 93*, 135–145.

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

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

Sweller, J. (1999). *Instructional design in technical areas.* Melbourne: ACER Press.

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 and selective attention as factors in the structuring of technical material. *Journal of Experimental Psychology: General, 119*, 176–192.

Sweller, J., & Cooper, G. A. (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 and Memory, 4*, 513–526.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 463–474.

Sweller, J., Mawer, R., & Ward, M. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112*, 634–656.

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Tarmizi, R., & 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.

Trafton, J. G., & Reiser, B. J. (1993). The contribution of studying examples and solving problems to skill acquisition. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 1017–1022). Hillsdale, NJ: Erlbaum.

Tuovinen, J., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334–341.

Van Merrienboer, J. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research, 6*, 265–287.

Van Merrienboer, J., & de Croock, M. (1992). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research, 8*, 365–394.

Van Merrienboer, J., & Krammer, H. (1987). Instructional strategies and tactics for the design of introductory computer programming courses in high school. *Instructional Science, 16*, 251–285.

Van Merrienboer, J., Schuurman, J., de Croock, M., & Paas, F. (2002). Redirecting learner's attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction, 12*, 11–37.

Vollmeyer, R., Burns, B., & Holyoak, K. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*, 75–100.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction, 7*, 1–39.

Yeung, A., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology, 23*, 1–21.