INSTITUTE OF COGNITIVE SCIENCE
Technical Report
University of Colorado, Boulder

Long-Term Working Memory

by

K. Anders Ericsson
The Florida State University
Department of Psychology
Tallahassee, FL 32306-1051

and

Walter Kintsch
Institute of Cognitive Science
University of Colorado at Boulder
Boulder, Colorado 80309-0344

ICS Technical Report #94-01

Long-Term Working Memory

by

K. Anders Ericsson
The Florida State University
Department of Psychology
Tallahassee, FL 32306-1051

and

Walter Kintsch
Institute of Cognitive Science
University of Colorado at Boulder
Boulder, Colorado 80309-0344

ICS Technical Report #94-01

Abstract

To account for the large demands on working memory during text comprehension and expert performance, the traditional models of working memory involving temporary storage must be extended to include working memory based on storage in long-term memory. In the proposed theoretical framework cognitive processes are viewed as a sequence of stable states representing end products of processing. In skilled activities, acquired memory skills allow these end products to be stored in long-term memory and kept directly accessible by means of retrieval cues in short-term memory, as proposed by skilled memory theory. These theoretical claims are supported by a review of evidence on memory in text comprehension and expert performance in such domains as mental calculation, medical diagnosis, and chess.

To perform complex cognitive tasks, people must maintain access to large amounts of information. For example, an individual reading a sentence in a text must have access to previously mentioned actors and objects to resolve references to pronouns. The individual also needs contextual information to integrate coherently information presented in the current sentence with the text previously read. Similarly, mental calculators must maintain the results of intermediate steps in memory, as, for example, when they mentally multiply two 5-digit numbers. Chess masters can play chess games without being able to see and manipulate a chess board, thus they are able to keep the current locations of all the chess pieces in accessible form in memory. These working contexts, with their accessible information that change as the individuals continue with the task, are often informally referred to as instances of working memory. The standard definition of working memory is more restrictive, however, and refers "to the *temporary storage* of information that is being processed in any of a range of cognitive tasks" (Baddeley, 1986, p. 34, italics added), that is, to information maintained in readily accessible storage for only a short period of time without rehearsal or re-activation. If the standard definition with its mechanism is accepted as an account of all instances of working memory several critical issues emerge. In subsequent sections we will focus, in particular, on the following: Can mechanisms that account for subjects' limited working-memory capacity in laboratory tasks also account for the greatly expanded working-memory capacity of experts and skilled performers? How can working memory based on temporary storage account for the fact that skilled activities can be interrupted and later resumed without major effects on performance?

In this paper we propose that a general account of working memory has to include in addition to the temporary storage of information that we refer to as short-term working memory (ST-WM), another mechanism based on skilled use of storage in long-term memory (LTM) that we refer to as long-term working memory (LT-WM). Information in LT-WM is stored in stable form, but *reliable access* to it may be maintained only temporarily by means of retrieval cues in ST-WM. Hence LT-WM is distinguished from ST-WM by the durability of the storage it provides and the need for sufficient retrieval cues in attention for access to information.

The classic distinction between STM and LTM (Atkinson & Shiffrin, 1968; Waugh & D. A. Norman, 1965) has remained a central feature of all major information-processing theories of memory (see Cowan, 1988, and Estes, 1988, for recent reviews). The new contribution we hope to explicate is that reliance on acquired memory skills will enable individuals to use LTM as an efficient extension of ST-WM in particular domains and activities after sufficient practice and training. Extending the Chase-Ericsson (1982) Skilled Memory Theory, we show that mechanisms similar to those underlying a ten-fold increase in performance on tests of STM are used by experts and skilled performers to extend their effective working-memory capacity. In particular, our extension will focus on mechanisms that allow skilled performers to overcome proactive interference caused by prior storage of similar information in LTM. According to our concept of LT-WM with reliable access, individuals rely on specific control processes to encode heeded information in retrievable form. Specifically, individuals draw on acquired knowledge and on systems of retrieval cues that we refer to as retrieval structures. Within our proposal for LT-WM we can easily account for skilled performers' expanded capacity of working memory in activities for which they have acquired knowledge and special memory skills. Furthermore, storage of information in LT-WM implies that most types of accessible information in working memory will remain in LTM during an interruption of skilled activities and can be easily reinstated by reactivation of necessary retrieval cues.

Our proposal that efficient storage and retrieval characteristics can be acquired for LTM by skilled performers might seem to blur the traditional functional distinctions between STM and LTM. Consequently, we will start the paper with a review of research on human memory to show that for skilled performance there is extensive evidence for flexibility of characteristics of LTM and the development of skills leads to : (1) increased speed of storage and retrieval on specific memory tasks through practice, (2) incidental storage of information in LTM during skilled activities and (3) memory skills for selective storage in and retrieval from LTM. We will then contrast our proposal to other theories of working memory. Unlike most other general theories of working memory, our concept of LT-WM is a memory skill that individuals acquire to meet the particular memory demands of a complex, cognitive activity in a particular domain. In separate sections on text comprehension and expert performance we will describe the mechanisms of LT-WM in detail.

A Brief Review of Research on Memory

Until quite recently researchers did not study directly the availability and accessibility of information stored in working memory during complex cognitive processes. Instead, they made inferences about working memory from studies of general memory capacity. Ever since Ebbinghaus (1885/1913) researchers have concentrated on deriving general laws and capacities for memory from simple tasks explicitly designed to study memory performance for arbitrary sequences of information. The standard procedure has been to present a list of unrelated items and to require reproduction with either immediate free recall or free recall after some interpolated activity. In the traditional model of human memory (Atkinson & Shiffrin, 1968; Waugh & D. A. Norman, 1968), immediate free recall yields items directly retrieved from a temporary short-term memory (STM) and items retrieved by retrieval cues from a more durable storage in long-term memory (LTM). STM is assumed to have a limited capacity of around seven chunks (G. A. Miller, 1956), a chunk corresponding to a familiar pattern already stored in LTM. Storage in STM is temporary, and when attention is diverted to another demanding task, information originally stored in STM becomes unavailable in a matter of seconds (Brown, 1958; L. R. Peterson & M. J. Peterson, 1959). In contrast, the storage capacity of LTM is assumed to be vast and much more durable than that of STM. Storage in LTM is assumed to be primarily associative, relating different items to one another and relating items to attributes of the current situation (current context). The time required for storage of a new retrievable memory trace in LTM has been estimated to be relatively long--about ten seconds (Simon, 1973). The primary bottleneck for retrieval from LTM is the scarcity of retrieval cues that are related by association to the desired item, stored in LTM.

During free recall subjects are asked to retrieve a list they studied earlier. At the time of recall the subjects have no direct retrieval cues for the list in STM, and many theorists (Walter Schneider & Detweiler, 1987; Shiffrin, 1976) have proposed that recall must be mediated by the cues available in the current context. If some of the cues in the context for recall were still available in the context for study of the list, they should be part of the memory trace and thus serve as retrieval cues for items in the list to be recalled. Once some items have been recalled, they can serve as additional retrieval cues. Consistent with this proposed mechanism, free recall in most list-learning experiments is relatively poor and decreases as a function of the number of lists the subjects have previously studied and thus linked to the same contextual cues.

In most of the prodigious research on memory, investigators have tried to isolate different memory systems and to specify their storage and retrieval characteristics with general measures that are independent of materials and of subjects' specific background knowledge. Researchers have focused on estimating maximal capacities by studying memory performance in simple tasks that test only memory. It is generally assumed that the same distinctions and capacity limitations observed for simple tasks apply to working memory in complex cognitive activities.

From General Theories of Memory to Models of Working Memory

General theories of memory, which are based on performance on memory tasks, incorporate constraints on storage and retrieval that are assumed to apply to any type of activity. When investigators began studying more complex cognitive processes, such as problem solving, decision making, and concept formation, the models they developed had to be consistent with these theories. An adequate model of performance in a task had to specify the relevant background knowledge and skills subjects had as well as sequences of processes that did not violate the constraints on the amount of information kept available in memory (that is, working memory).

"Availability of information" generally implies both rapid and reliable, essentially error-free, storage of and access to that information. In traditional memory experiments these criteria for availability are met only for information stored in STM. Because our proposal for working memory is based on storage in LTM, we briefly review some of the major reasons that LTM has been considered insufficient to support reliable storage of available information.

In standard theories of memory (Atkinson & Shiffrin, 1968) information can be stored in LTM only after it has been stored in STM, and even then, storage in LTM is a probabilistic event. Originally, Atkinson and Shiffrin proposed that the probability of storage in LTM is a function of the time an item was maintained in STM. More recently, Anderson (1983) suggested that the probability of storage is a function of the number of times an item enters STM. Subjects' control of the storage of information appears to be limited, as shown, for example, by low levels of free recall in list learning. Furthermore, in more meaningful tasks subjects' recall of presented information is not improved when they are instructed to study that information for later recall (Craik & Lockhart, 1972). This finding implies that subjects cannot achieve reliable storage of information in many of the standard memory tasks. Anderson (1983) goes even so far as to argue that subjects' inability to control storage in LTM is beneficial since they cannot predict what information will be useful later on. The memory performance exhibited by subjects in standard memory tasks is clearly consistent with the view that storage of information in LTM and efficient access of that information is too unreliable to be an effective source of working memory. We argue later in this paper, however, that the performance of untrained subjects who memorize lists of unrelated items in the laboratory does not accurately describe the efficient storage and retrieval that experts in specific domains can achieve after many years of practice.

Models of working memory have focused on the availability of information in STM which has limited capacity. No model of working memory can reasonably allow greater working capacity during performance of a specific task than the maximal capacity of working memory measured in a pure memory task. That is, the capacity of working memory must be much less than the memory span of seven chunks (G. A. Miller, 1956), in which perfect recall is achieved only 50% of time on the average. Hence the capacity of reliable working memory is assumed to be only around four chunks (Broadbent, 1975). Such a severe limit on working memory might seem far too restrictive to allow for human performance levels. Laboratory studies have shown, however, that a wide range of cognitive activities, such as problem solving (Atwood, Masson, & Polson, 1980; Newell & Simon, 1972), concept formation (Levine, 1966), and decision making (Payne, 1976), can be successfully accounted for by models that permit storage of a very small number of products in STM.

Newell and Simon (1972) proposed a production-system architecture for cognitive processes that has influenced most subsequent efforts to build models and theories. In this architecture the conditions of a large number of productions (condition-action pairs) are matched against the currently active elements (working memory). In more recent models, such as Anderson's (1983) ACT*, working memory is the transiently activated portion of LTM. The limits on the number of elements in working memory are not determined by a fixed number but rather by the amount of available activation. In his work on building ACT* models of cognitive processing Anderson found that working memory can sometimes contain over 20 units at one time. To reconcile such a large capacity of working memory with the much smaller capacity of STM, Anderson (1983) argued as follows: The activation of elements decays very rapidly. For this reason the number of units that can be actively maintained long enough to be included in immediate recall is much less than all of the information activated at the start of recall. Most investigators argue, however, that the capacity of working memory must be far greater than the capacity of traditional STM (Newell, 1990).

Working memory in production-system architectures was originally viewed as a single construct with general resources. In a very influential line of empirical research initiated by Baddeley and Hitch (1974), investigators examined this assertion by studying the effect on cognitive performance from an additional concurrent task specifically designed to interfere with the capacity of working memory. The result of over a decade's active research on that and related paradigms (reviewed by Baddeley, 1986) conflicted with the single-construct view. Although reliable decrements in speed and/or accuracy of cognitive processes were often obtained for the main task when an additional memory task was introduced, the primary task performance decreased for the most part only slightly even when subjects had to maintain 3 to 6 digits (near their digit span) in working memory while executing the primary task. To account for these findings Baddeley (1986) proposed that in addition to a central executive there are two slave systems, the articulatory loop and the visuo-spatial scratch pad, in which the central executive can store information temporarily. Investigators have obtained converging evidence for these subsystems by examining the relation between individual differences on the main task and on tasks measuring various types of memory performance. Of particular interest are findings from neuro-psychological patients who exhibit exceptionally poor performance on tasks that measure the capacity of one of the subsystems, for example, memory span for words which is assumed to measure the capacity of the articulatory loop. Consistent with the independence of the subsystems in Baddeley's model, patients with dramatically impaired subsystems are still able to perform text comprehension tasks at normal levels (Baddeley, Papagno & Valla, 1988; Baddeley & Wilson, 1988; Martin, 1993). At the same time this finding means that working memory in such skilled activities as text comprehension must be accounted for by the central executive and thus remains essentially unexplained.

In sum, recent research has shown that working memory does not consist of a single general capacity, but rather consists of several subsystems that can be relied on to complete various types of tasks (Baddeley, 1986). Interference with either of these subsystems by a concurrent secondary task degrades performance on the primary task only slightly. This finding implies that the central executive has sufficient working memory capacity to complete the processing. Of particular interest to the LT-WM we propose is the observation that concurrent memory tasks appear to cause the least impairment in highly skilled activities such as piano playing (Allport, Antonis & Reynolds, 1972), typing (Shaffer, 1975), and reading (see Baddeley, 1986, for a review). As explained in the following section, general theories of memory provide a better account of working memory when subjects are performing unfamiliar tasks than they do when subjects are performing skilled activities, for which demands on working memory are considerably larger. This finding is consistent with our hypothesis that working memory in a specific, skilled activity increases as one aspect of acquired skill.

Differences Between Memory in Standard Memory Tasks and in Complex Cognitive Performance

The cognitive skills individuals apply to complex tasks are fundamentally different from those they apply to standard memory tasks with arbitrary materials. We note three significant differences and demonstrate how, within a specific domain, individuals can acquire memory skills and thereby increase the available working memory capacity beyond the domain-general capacity proposed by traditional theories of memory. First, individuals attain a particular level of performance in complex cognitive tasks only over extended periods of time, and much of the information is familiar and meaningful to their activity. In contrast, traditional studies of memory minimize relevant experience by using naive subjects and lists of unrelated items. As part of the acquisition of skill, individuals acquire the ability to rapidly store in LTM relevant information which they are presented. Contrary to earlier views, it is possible for individuals to reliably store selected information in LTM even when presentation rates are relatively fast. Secondly, unlike the directed and effortful task of memorizing as much as possible from a list, storage of relevant information in LTM during skilled performance is a natural consequence of performing. The ability to select relevant information and anticipate future retrieval demands is an integral aspect of acquired skill and reduces dramatically the amount of information that must be stored. Finally, skilled performance requires that all relevant information be rapidly and reliably accessible. Traditional memory paradigms involving recognition tests and free recall were developed to assess whether some input was stored in memory rather than whether some particular task-relevant information could be accessed reliably within a few seconds. Rapid recall of some information was interpreted as evidence that information had been stored in STM. A different methodology for measuring the speed and accuracy of selective recall is required for assessing the accessibility of information in working memory.

Increased Memory Performance due to Acquired Knowledge and Skill

It is generally observed that memory performance increases after an individual practices on memory tasks involving specific types of materials and that an individual's familiarity with a given type of material is related to the amount of material recalled (see Ericsson, 1985, for a review). Increases in memory performance are not in themselves inconsistent with the notion that the capacity of the human memory system is invariant, and they have traditionally been accounted for in terms of chunking (G. A. Miller, 1956). With more familiarity and experience with a particular type of stimulus material, subjects acquire over time a set of complex patterns in LTM that allows them to represent subsequently presented information in terms of already acquired patterns (chunks) of elements rather than individual stimulus elements. The most direct predictions of the chunking model can be made for rapidly presented information, where storage in LTM is minimized or even be completely eliminated. Under those conditions, recall of presented information is assumed to be restricted to the activated chunks in STM. Consistent with these predictions subjects can recall about the same number of random letters (consonants) as unrelated words, where each word corresponds to a chunk of letters (G. A. Miller, 1956).

Large differences between experts' memory performance and that of novices have similarly been accounted for in terms of chunking. In their pioneering research on the superior memory of chess experts, Chase and Simon (1973) proposed that after many years of study chess experts have stored a large number of specific patterns of chess pieces (chunks) in LTM. This memory representation allows them to rapidly recognize several patterns in a presented chess position and thus to encode and recall many chess pieces by relying only on the fixed number of chunks in STM. Consistent with the hypothesis that the chess experts' superior memory was mediated by familiar meaningful configurations of chess pieces, Chase and Simon found that the experts' advantage disappeared when chess boards of randomly arranged chess pieces were used as stimuli in the memory tasks. Experts' superior memory for representative stimuli from their domain of expertise, but not for randomly rearranged versions of those stimuli, has been frequently replicated in chess (see Charness, 1991, for a review) and also demonstrated in bridge (Charness, 1979; Engle & Bukstel, 1978); go (Reitman, 1976); music (Sloboda, 1976); electronics (Egan & Schwartz, 1979); computer programming (McKeithen, Reitman, Rueter, & Hirtle, 1981); dance, basketball, and field hockey (Allard & Starkes, 1991); and figure skating (Deakin & Allard, 1991).

Consistent with chunking theory Chase and Simon(1973) proposed that expert memory performance could be accounted for solely in terms of STM, where chunks were temporarily kept or activated. Charness (1976) found, however, that information about chess positions was indeed stored in LTM. When other tasks were interpolated to eliminate any information about the chess position in STM, no or minimal effect on subsequent recall performance was observed. Furthermore, Frey and Adesman (1976) found that the hypothesis of a fixed number of chunks in STM was untenable for chess experts' memory of chess positions. When two chess positions were sequentially presented, their subjects recalled either one almost as well as they did after seeing only a single position. Although Frey and Adesman found some evidence for intrusion errors between the two positions, distinct memory for the two positions in LTM was demonstrated. More recently, Cooke, Atlas, Lane and Berger (1993) extended these findings and showed that highly skilled chessplayers can recall substantial amounts from up to nine different chess positions, which had been presented in sequence at a comparably fast rate of presentation. In two additional experiments Cooke et al. demonstrated that skilled chessplayers encode chess positions in terms of high-level descriptions of their structure that allow the players to distinguish memory encodings of multiple chess positions from each other.

The largest individual differences in memory performance are associated with memory experts and professional mnemonists, whose exceptional performance reflects high acquired abilities to store specific types of information in LTM (see Ericsson, 1985, 1988) for a review). Some of the most detailed information on the structure of acquired memory skill has been collected in studies of the effects of extensive practice on the digit-span task. In this task, which was designed to measure the capacity of STM, subjects strive for complete serial recall of a rapid sequential presentation of digits. After hundreds of hours of practice on this task two subjects were able to increase their memory performance from around 7 digits, which is the typical performance of untrained subjects, to over 80 digits (Chase & Ericsson, 1982; Staszewski, 1988a). Other subjects have acquired digit spans of over 20 digits within 50 hours of practice (Chase & Ericsson, 1981; Ericsson, 1988a). Detailed experimental analyses of the superior recall performance of these trained subjects show that their performance reflects storage in LTM and is limited to the specific type of material practiced. Along with increases in memory performance, the speed with which subjects can memorize a list of a given length increases and can match the presentation rates in memory tasks designed to study only STM. To attain accurate recall from LTM, subjects associate the presented items with predetermined retrieval cues (retrieval structures) during the initial presentation and then activate these retrieval cues to retrieve the items during the subsequent recall phase.

In summary, recent research on memory performance shows that with practice and the acquisition of memory skills, subjects can improve their recall performance on a specific memory task with a particular type of stimulus material by 100 to 1000%. Their virtually perfect reproduction of presented information, especially in the digit-span task, demonstrates that after practice subjects are able to use LTM for reliable storage even at relatively fast presentation rates. The improvement is due to increased ability to store information in LTM and to the association of presented information with retrieval cues that allow reliable retrieval from LTM at the time of recall.

Storage in LTM during Task-Oriented Processing

The evidence reviewed above shows that subjects can rapidly store information in retrievable form in LTM under constrained conditions. However, demonstrations of this capacity in tasks focused solely on memory performance does not prove that the same or similar mechanisms can be and are used for storage to extend working memory during skilled cognitive activities. When subjects perform a skilled cognitive activity, they focus on the goals of that activity. It is thus possible to assess storage in LTM of information relevant to that activity by stopping subjects unexpectedly and asking them to recall specific information about the preceding activity. Traditionally investigators have concentrated on incidental memory of tasks and other externally presented information, for which accuracy of recall and recognition is easy to score. Because we are interested in the detailed mechanisms of working memory, however, we should ideally analyze the activity under study so that we can predict what is relevant in the presented information and what intermediate products would need to be stored in WM for successful execution of the task.

A fairly large body of research using free-recall tests has assessed incidental memory of stimuli encountered by subjects who were completing different kinds of tasks. This research has mostly used relatively simple tasks, such as evaluations and judgments, which can be factorially combined with different stimuli. The most intriguing result was that incidental memory in some judgment tasks was as good as the memory of subjects who deliberately memorized the same stimuli (Craik & Lockhart, 1972; Hyde & Jenkins, 1973). Furthermore, subjects' memory for the presented stimuli did not benefit from their knowing about a subsequent recall test in advance. Hence memory of the stimuli must be a direct consequence of the cognitive process generating the judgment. In their influential theoretical framework of depth of processing Craik and Lockhart (1972) describes a general account of memory in terms of different encoding operations within an activity. The retrieval operations should duplicate the encoding conditions in order to be effective (Moscovitch & Craik, 1976) in accord with the encoding-specificity principle (Tulving, 1979).

Investigators have often used more complex tasks with greater demands on working memory, such as categorization of a large number of stimuli into several different categories. They have found high levels of incidental free recall that relate directly to the number and size of the categories of stimuli that subjects generate (Mandler, 1967). Research on expertise shows that after engaging in typical activities in their domains experts have substantial incidental memory for information relevant to their particular tasks. In his pioneering research on chess expertise, de Groot (1946/1978) found that chess masters after selecting the next move for an unfamiliar chess position could reproduce virtually the entire position as part of their verbal reports on their thought processes during the selection of the move. A subsequent experimental study by Lane and Robertson (1979) showed that chess players' memory of a chess position after the move-selection task is just as good as it is when the players have been informed about the subsequent memory test in advance. Moreover, the amount of correctly recalled information is correlated with the level of subjects' chess expertise (Charness, 1981a; Lane & Robertson, 1979). Mere exposure to the chess position is not sufficient to produce a good memory of it. When chess players engage in activities with a chess position that are unrelated to chess, such as counting chess pieces whose color matches that of the corresponding square, their incidental memory is minimal and prior knowledge of a subsequent memory test leads to large increases in recall performance (Lane & Robertson, 1979). Unexpected recall of the configuration of cards in bridge hands yields a very similar pattern of results, and the amount of accurate incidental recall increases as a function of expertise in bridge (Charness, 1979; Engle & Bukstel, 1978). Incidental memory for data on patients after a medical review and diagnosis is far greater for medical experts than for medical students (G. R. Norman, Brooks & S. W. Allen, 1989). However, when the same information about patients is presented to medical experts and students who have been informed that memory test will be given afterwards, the difference in their recall performance is dramatically reduced and no longer reliable.

In sum, complex processing of stimuli leads to substantial incidental memory that subjects can successfully retrieve at the time of subsequent free recall. These results imply that a lot of relevant information is stored in LTM during the normal skilled processing for a task and that this information must have been potentially accessible during the work on that task. Free recall of information after the cognitive activity is completed does not necessarily provide an accurate estimate of the amount of information individuals can efficiently access during processing (working memory). Information in

LTM can be retrieved only with the relevant cues, and not all the cues available during some cognitive activity are also spontaneously available in a free recall task (cf. Tulving & Pearlstone, 1966).

Selective Retrieval of Relevant Information from STM and LTM

The distinctive functional characteristic of working memory is that from it, individuals can selectively access both relevant presented information and intermediate results they have already generated. The traditional concept of STM includes both of these types of information and hence provides an appealing account for the entire capacity of working memory. Because only a limited number of units (chunks) are activated at a given time and are therefore distinguishable from the rest of the vast amount of information stored in LTM, access of information from this small set of information in STM is not problematic. Furthermore, highly activated information is consistent with rapid and efficient access. A brief review of some of the estimated characteristics of access from STM is appropriate here because these characteristics can be used as a reference point for evaluating our proposal for access from LT-WM.

Most studies of STM have not studied representative types of retrieval from working memory but rather focused on measuring the fundamental parameters of this memory store. Speed of access is most commonly measured by recognition tests. After subjects have had some limited practice with this type of memory task (Kristofferson, 1972) information stored in STM is very rapidly recognized with search rates estimated at 25-30 items/s (Sternberg, 1969). However, the speed of recognition cannot be equated with the speed of retrieval of desired information because no new information is selectively brought into attention. The times for recognition should be viewed as lower bounds for selective retrieval.

Selective retrieval can best be studied with cued recall. In cued recall tasks, subjects retrieve an item in response to a retrieval cue provided by the experimenter. Retrieval from STM is usually elicited by the experimenter showing the subject a list of items and later asking the subject to name the item that followed a given item in the list. Sternberg (1969) found that retrieval time was a linear function of the length of the list. This result implies a search rate of 250 ms per item, which is almost ten times slower than the rate assessed for recognition. Several investigators (Weber & Blagowsky, 1970; Weber & Castleman, 1969; Weber, Cross & Carleton, 1968) have attained similar estimates for retrieval from STM. Weber and his colleagues have noted that the search rates are very close to the speed of internal speech (Weber & Bach, 1969). According to chunking theory not all information about the chunks is stored in STM. Instead a general pointer or retrieval cue is stored in STM that allows the subject at the time of recall to access information from the chunk in LTM. The time necessary to retrieve all the elements of a single chunk was estimated by Yu, W. Zhang, Jing, Peng, G. Zhang and Simon (1985) to range from 281 ms to 373 ms. Hence, selective retrieval of information "stored in STM" is far from immediate, and combined search and retrieval might require a duration of around a second.

The time it takes to access information stored in LTM is generally estimated from the difference between recognition times for items previously stored in LTM and for just seen items that are retained in active form in STM. As noted earlier, storage in LTM is not generally immediate for typical laboratory stimuli, and for lists of unrelated items investigators provide subjects with study times sufficient for storage in LTM prior to a test of recognition. Sternberg (1969) found that recognition times for highly practiced lists stored in LTM were 200 ms to 400 ms longer than those for items stored in STM, depending on the length of the list (up to five items long). For meaningful materials such as sentences, storage in LTM occurs typically after a single presentation. In recognition tests of sentences Anderson (1990) found that access to sentences stored in LTM takes somewhat longer than it does for just seen sentences stored in STM. An additional 420 ms is required for sentences stored after a single presentation. For sentences studied at two different times the estimated retrieval time is reduced to 280 ms (Anderson, 1990).

Measurement of retrieval speed in studies of cued recall of information in LTM are remarkably rare. When unrelated paired associates such as foreign vocabulary items, are learned subjects need a fair amount of study time. Even after considerable study time cued-retrieval times are slow--typically around 2 s (Crutcher, 1990)--and only after extensive retrieval practice are access times below 1 s (Crutcher, 1992). With more familiar materials, such as paired associates of one-syllable English words Waugh (1985) found that retrieval times were around 800 ms after the first study trial and decreased to around 400 ms after 18 practice blocks, which was indistinguishable from the times taken to read the stimulus words aloud.

As we stated earlier, the use of LTM as an extension of working memory is possible only under very restricted circumstances. Subjects must be able to rapidly store information in LTM, an ability that requires extensive experience and a large body of relevant knowledge and patterns for the particular type of information involved. Our review showed that such abilities are primarily observed in memory experts and other types of experts for specific domains. Furthermore the activity has to be very familiar to the experts because only then can they accurately anticipate future retrieval demands for information. These conditions allow selective storage of information. Finally subjects associate the encoded information with appropriate retrieval cues, which allow them at a later time to activate a particular retrieval cue thus partially re-instating the conditions of encoding to retrieve the desired information from LTM. When a set of retrieval cues are organized in a stable structure we will refer to that structure as a retrieval structure. The acquired memory skill involves the development of encodings for which the subject can provide controlled access to significant aspects of the encoding context and thus indirectly to the desired information in manner consistent with the encoding-specificity principle (Tulving, 1983). Hence the particular instanciation of LT-WM in a given activity will correspond to a domain-specific memory skill that will differ considerably in its specific structure from memory skills in other activities. The difficulty of examining LT-WM in detail during experts' performance of representative tasks in their domain of expertise is due to the complexity of the pre-requisite knowledge and the corresponding tasks. Later in this paper we discuss the structure of LT-WM during text comprehension and in different forms of expert performance in various domains.

At a very general level we can characterize LT-WM as mediated by a retrieval schema, in which information the subject has encountered is encoded and stored in LTM, where it is associated with its appropriate retrieval cues as illustrated in Figure 1.

Insert Figure 1 about here

At the time of selective recall only the node corresponding to this specific structure needs to be available in STM, along with the retrieval cue specifying the type of desired information. For example, a medical doctor reviewing a chart for a patient selectively encodes that patient's test results in LT-WM to allow retrieval of that information from memory when that test result is relevant and specified by a corresponding retrieval cue.

Most of our current knowledge about LT-WM came originally from studies of exceptional recall of digits. In this task the demands for future retrieval are unusually clear: subjects are asked to reproduce all the digits in the exact order presented. Based on an intensive study of the encoding processes used by a subject, SF, who through training had acquired an exceptional digit span, Chase and Ericsson (1981) were able to assess a hierarchical retrieval organization that SF used to memorize lists of 30 digits shown in Figure 2. At the lowest level SF used mnemonic association to running times and other numerical relations to group digits and encode them as units. SF then used spatial relations to encode digit groups into super-groups.

Insert Figure 2 about here

At the time of recall, SF could easily regenerate any of the unique locations of the retrieval structure and use a given location as a cue to retrieve the corresponding digit group.

By proceeding sequentially through a retrieval structure, subjects can serially recall all of the digits in their presented order. Directly relevant to the characteristics of working memory, Chase and Ericsson (1981) found that the digits groups were accessible when descriptions of their location within the retrieval structure were used as cues. Chase and Ericsson demonstrated this form of accessibility experimentally in a cued recall task. After the subject had memorized a digit sequence, the experimenter presented locations in the retrieval structure and asked the subject to recall the corresponding digit group or asked the subject to point to the location of a presented digit group. With more extensive practice on the digit-span task the same type of retrieval by a second subject (DD) has been found to be virtually immediate (Staszewski, 1988a).

Figure 1

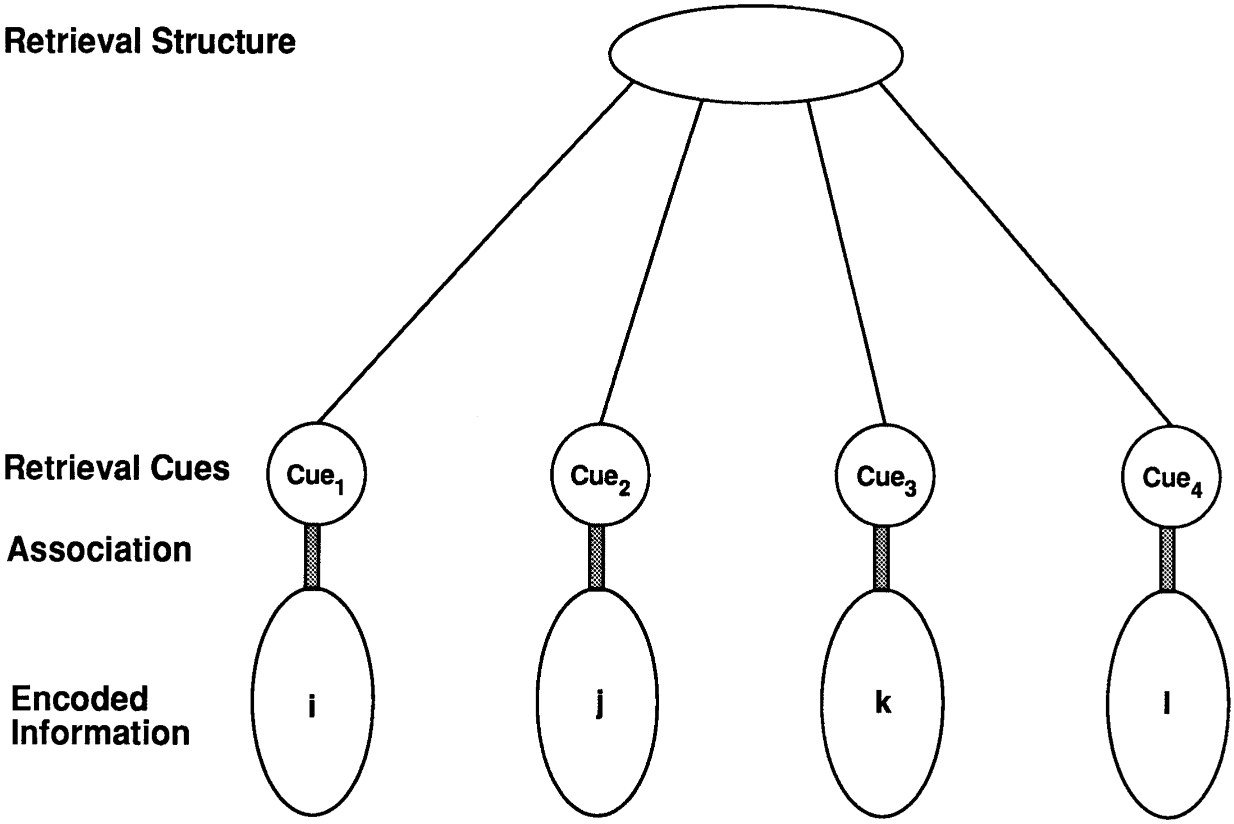
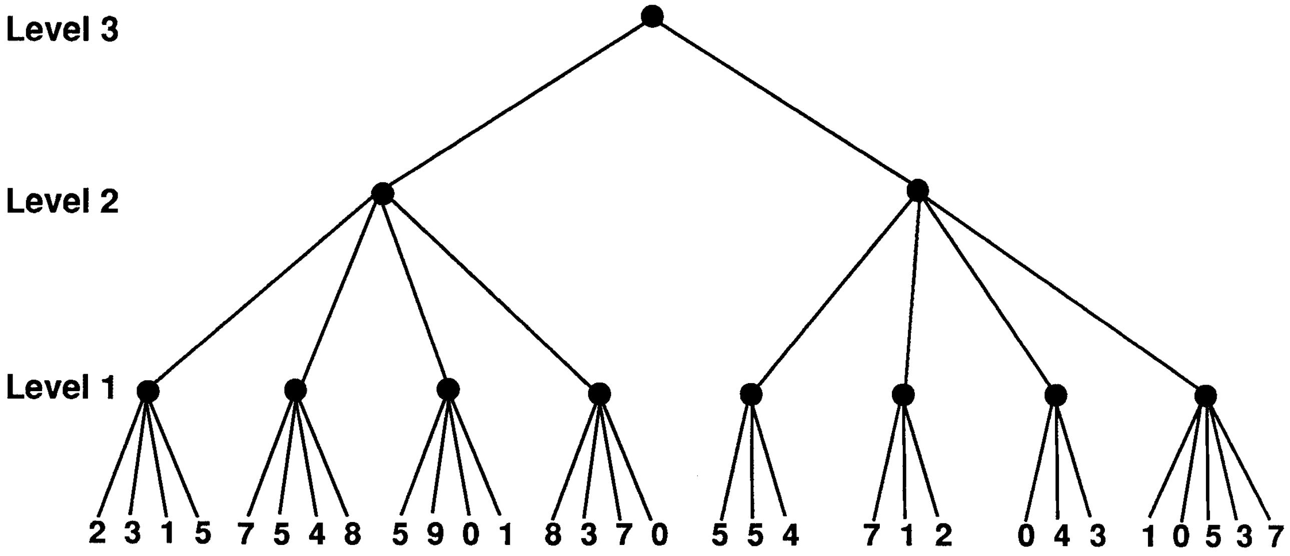

Retrieval Structure
Retrieval Cues
Association
Encoded Information
Cue$_1$
Cue$_2$
Cue$_3$
Cue$_4$
i
j
k
l

Figure 2



Further evidence for this immediate and flexible retrieval through retrieval structures is shown by these trained subjects' ability to memorize matrices of digits without any additional practice, and in particular to retrieve those digits in many different recall orders, as illustrated in Figure 3.

Insert Figure 3 about here

The trained subjects memorized the matrix in terms of five digit groups, each digit group corresponding to a different row. The speed of these subjects' storage and flexible recall matched that of the exceptional subjects studied by Binet (1894) and Luria (1968) and thus meets the criteria specified by these investigators for exceptional visual/photographic memory. However, both trained and exceptional subjects' times to recall the matrix were found to be a linear function of the number of times they had to retrieve a different digit group (row) to complete recall according to the specified order (Ericsson & Chase, 1982). The time required to recall a new digit group was estimated to be about 2 s, but a subject familiar with this particular recall task completed the retrieval within 1 s (Müller, 1917). Other studies of memory experts and subjects with alleged exceptional memory recall (reviewed by Ericsson 1985, 1988a) almost always yield evidence for the explicit use of retrieval structures in exceptional serial recall. Often the retrieval structures are hierarchies of spatial locations, but frequently subjects rely on a sequence of locations, as in the method of loci (Yates, 1966).

In sum, we argue that selective retrieval of information stored in LTM after a brief single presentation can be achieved with appropriate retrieval cues at speeds comparable to those for retrieval from STM. Hence working memory based on storage in and retrieval from LTM could attain through practice storage and retrieval speeds similar to those for STM.

The Chase-Ericsson (1982) proposal for Skilled Memory has been generally accepted as a mechanism for accounting for exceptional memory (Anderson, 1990; Baddeley, 1990; Carpenter and Just, 1989; Newell, 1990; Walter Schneider & Detweiler, 1987; van Lehn, 1989), but several investigators have voiced doubts about its generalizability to working memory. Carpenter and Just (1989) wrote that "Memorizing a sequence of digits this way shares some properties of language comprehension, namely the circumvention of working-memory limitations by on-line binding of successive elements to a hierarchical knowledge structure." (p.54). However, they argued that, unlike sentence comprehension, "the knowledge structure to which the digits are bound is fairly inflexible and typically known before hand" (p. 54). Walter Schneider and Detweiler (1987) argued that "it is important not to use LTM as working memory. This is because the faster LTM is changed, the greater the likelihood that retroactive interference will distort all the previously stored LTM, making it useless" (p.84). In support of the claim that working memory of experts does not rely on LTM, Baddeley (1990) interprets the finding that expert abacus calculators are unable to recall more than one sequence of presented digits (Hatano & Osawa, 1983) as evidence against any mediation of storage in LTM for their exceptionally large working memory capacity for numbers. In the next section we will try to show that Chase-Ericsson's Skilled Memory Theory can be extended into our conception of LT-WM that successfully addresses these concerns.

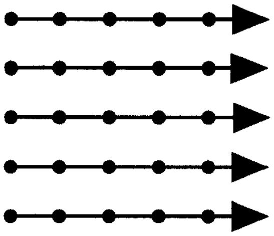
Proactive Interference in Long-Term Working Memory

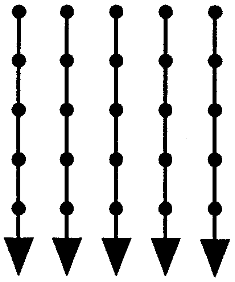
Efficient LT-WM requires that the skilled performer be able to anticipate future retrieval demands and thus identify and refine retrieval cues that uniquely index selected information. By reactivating the corresponding retrieval cue the skilled performer can access the desired information in memory. This mechanism should provide controlled retrieval as long as a given retrieval cue is associated with only a single piece of information. In many skilled activities, however, results generated during a task are frequently transformed and modified, or a series of similar tasks are performed within a relatively short time period. Hence the same retrieval cue is linked to several different results in succession. In this case the retrieval cue should access only the most recently stored result.

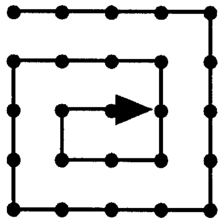
Theories of LTM predict considerable proactive interference from previously stored results and therefore a small likelihood of reliable retrieval from LT-WM under such circumstances. Inaccessibility due to interference is also postulated as an important cause of forgetting in LT-WM. In spite of these well-known findings our proposal for LT-WM is sometimes incorrectly believed to imply extended accessibility to all results stored in LTM regardless of intermediate activities. Hence a

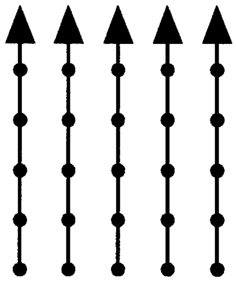
Figure 3

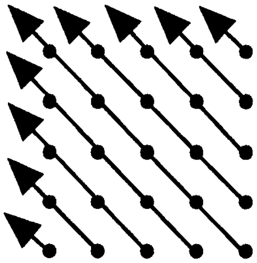
4	7	1	0	2
3	0	4	3	6
2	1	1	4	8
8	7	4	2	9
1	5	2	7	9

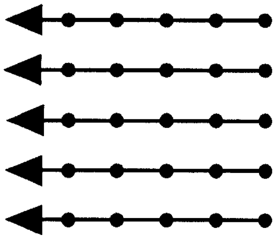

ROWS


COLUMNS


SPIRAL


UPWARD COLUMNS


DIAGONAL


BACKWARD ROWS

plausible model of LT-WM needs to demonstrate how the effects of interference can be overcome by distinct temporal separation and supplementary memory encodings.

It is well established that subjects can repeatedly use the same retrieval structure, such as the method of loci or pegwords, to encode successive lists of words and still be able to retrieve the most recently studied items (Bellezza, 1982). In a recent review Baddeley and Hitch (1993) showed that superior recall of the most recent experiences of some type is a very general phenomenon and has been successfully related to the temporal separation of the experiences over time periods ranging from seconds to months. As long as the temporal separation between the most recent encoding and the previous encodings is sufficient to make it distinctive, retrieval is accurate.

Temporal separation may also be important for explaining a finding from the Brown-Peterson paradigm that has been very difficult to explain by traditional theories of STM (Crowder, 1993). On the first trial with the Brown-Peterson paradigm, subjects' recall of presented information is quite accurate and does not decay as a function of the length of time they spend on the interpolated activity. Only after three or four repeated trials has sufficient proactive interference been accumulated to induce the typical decay of recall that occurs as a function of the time spent on the distracting activity. Walter Schneider and Detweiler (1987) showed that temporal distinctiveness (recency) could account for this phenomenon. They also found in a review of earlier studies that long intertrial intervals of around 2 m between successive Brown-Peterson trials essentially eliminate interference and restore recall to the level of the first trial of the session. If we assume that the presented information can be encoded by associations to corresponding category cues, so that the category cues serve as retrieval cues for subsequent recall, we can extend the recency mechanism to explain other related findings involving presented instances of distinct categories. We can account for the release of proactive interference due to an experimental change in the category of presented instances (Wickens, Born & C. K. Allen, 1963) and also for multiple recency effects for different categories after presentation of mixed lists (Watkins & Peynircioglu, 1983).

The preferred laboratory paradigm for studying interference involves presenting paired associates, where the stimulus item of the associate is recombined (updated) with different response items. Within this paradigm recall for the most recently presented response is typically measured using the stimulus item as a cue, but it is also possible to measure final recall of all the presented responses for a given stimulus item. The pattern of performance on these two recall tasks is dramatically influenced by the subjects' encoding methods (Bjork, 1978). When subjects memorize the paired associates by ordered rehearsal, the typical pattern of interference is obtained: cued recall of the most recent responses is accurate but old responses couldn't be recalled at the final memory test (study by Bjork and McClure reported in Bjork, 1978). Other subjects in the same experiment were instructed to memorize responses to the same stimulus item by constructing a story the interrelated all the responses in the order presented. Cued recall of the most recent responses for this group of subjects matched that of the group using ordered rehearsal, but final recall of all responses was much higher and close to the level of recall for the most recent responses. Bjork (1978) summarized a large body of research and concluded that all methods of encoding associated with updating have positive and negative features and only in the context of the retrieval demands for a particular task can the best method be selected.

In many cases it is important to overcome interference form several different items associated with the same concept or retrieval cue by generating an integrated memory representation including all items. One effective method to achieve integration for unrelated words is the story-construction mnemonic discussed above. This method has a long history in memory research and is effective to memorize a list of words or store many items in the same physical location in the method of loci (See Bower, 1972, for a review). A similar type of integration is now also established in one of the dominant paradigms for studying the effects of interference in LTM--the fan effect for recognition of studied facts. In this original paradigm Anderson (1974) had subjects memorize a series of sentences. Each sentence has a subject placed in a location, such as "The hippie is in the park", "The lawyer is in the park", and "The hippie is in the bank". The key finding was that the time to correctly recognize a given sentence increased as a function of the number of occurrences (fan) of that sentence's subject and location descriptions among the other sentences. Although the fan effect is highly reproducible with facts memorized independently, it is reduced (Moeser, 1979; E. E. Smith, Adams & Schorr, 1978) or even reversed (Myers, O'Brien, Balota & Toyofuku, 1984) for a series of sentences that forms an integrated representation, such as an episode of christening a ship. When subjects memorize sentences for multiple themes, a fan effect for the number of different themes emerges, but the number of irrelevant facts does not have a reliable influence (Reder & Ross, 1983). Pre-experimental information used to generate the integrated encodings is proposed to be a major factor mediating the increased recognition performance for integrated facts (Jones & Anderson, 1987). Recently, Radvansky and Zacks (1991) have found an intriguing asymmetry for sentences describing the location of objects. Recognition times for a sentence shows no fan effect for the number of other objects in the same location, whereas the number of locations associated with a given object leads to a typical fan. They propose that it is possible to store objects in the same location in an integrated memory representation similar to a mental model (Johnson-Laird, 1983), and we would add a situation model (van Dijk & Kintsch, 1983). It is doesn't appear possible to integrate several different locations associated with the same object in a similar manner, at least not for arbitrary spatial locations. In our subsequent section on text comprehension we will show that the generation of integrated memory representations is the normal mode of processing texts on familiar topics.

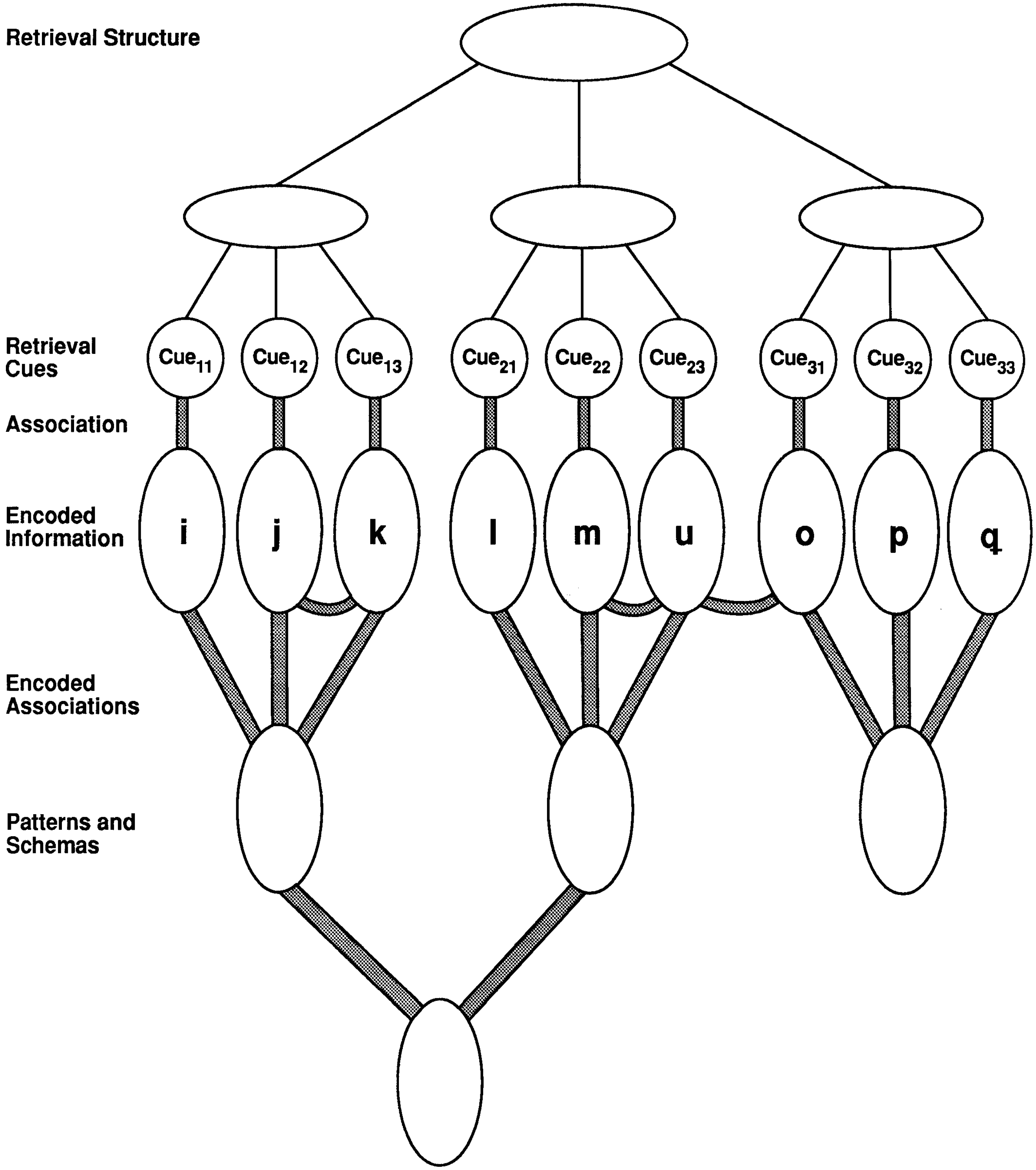
Comparatively little attention has been directed toward the effects of interference in working memory of skilled performers. Chase and Ericsson (1982) found that the performance of their trained memory experts was influenced by interference in several ways. In their procedure for testing digit span, consecutive digit lists had nearly the same number of digits and thus required the use of the same retrieval structure. On the first memory trial of a session, when proactive interference from other memorized digit lists was minimal, our subject was more likely to recall the digit list perfectly than he was on subsequent memory trials, for which the probability of perfect recall was considerably lower. However, the decrement in the total number of correctly recalled digits was small, a result suggesting that inference has a reliable, yet limited, influence.

The influence of interference is more dramatic during post-session recall of previously presented lists. Chase and Ericsson (1981, 1982) found that although their trained subjects were able to recall the majority of all presented 3- and 4-digit groups during post-session recall, they were typically unable to recall the exact sequence of entire lists of presented digits except for the last one or two lists. For the post-session recall the subjects used the mnemonic categories of running times as retrieval cues rather than the retrieval structure itself. Hence proactive interference restricted the retrievability of previously presented lists through retrieval structure cues. In a direct test of the effects of maximal interference, Chase and Ericsson (1982) tested DD on a paradigm related to one used by Frey and Adesman (1976) for recall of chess positions. First, one list of digits was presented. After a brief pause another list of the same length was presented and immediately recalled. Only then was the first list recalled. Under those conditions DD recalled the second list virtually perfectly, but was also able to recall most of the first list correctly. Recall accuracy for the first list ranged from virtually perfect to around 70%, which incidentally is close to the decrement originally observed by Frey and Adesman (1976) for chess experts' recall of chess positions. From concurrent reports that the trained subjects gave during memorization and recall of the digits, Chase and Ericsson determined that these subjects encode not only the digit groups together with their retrieval cues, but also higher-level relations between mnemonic categories within and between supergroups. For example, the first four 4-digit groups in the first supergroup (see Figure 2) was described in a retrospective report by one of the trained subjects as " um, the whole first four groups of four, it just went two ages, mile, mile two ages, and the miles were similar and the two ages were similar, so I was just set on that (Staszewski, 1990, p. 259)". Hence the trained subject encoded patterns of mnemonic codes to build a structure in LTM of the digit sequences (see Chase & Ericsson, 1982, and Staszewski, 1990, for a review of the extensive empirical evidence). Figure 1 must thus be augmented with additional encodings that generate associations forming a new memory structure connected to most of the digits in the list, as shown in Figure 4.

Insert Figure 4 about here

In addition to the direct associations between the encoded information and retrieval cues, subjects build a unique structure in LTM where the elements are directly related by semantic relations. From these findings we hypothesize that this generated structure will be relatively immune to proactive interference. Moreover, an element in this structure combined with cues in the retrieval structure should allow constrained recall of information belonging to this specific structure.

Figure 4



Retrieval Structure
Retrieval Cues
Association
Encoded Information
Encoded Associations
Patterns and Schemas
Cue11
Cue12
Cue13
Cue21
Cue22
Cue23
Cue31
Cue32
Cue33
i
j
k
l
m
u
o
p
q

In sum, our proposal for LT-WM includes cue-based retrieval without additional encodings (see Figure 1) and cue-based retrieval with an elaborated structure associating items from a given trial or context (see Figure 4). The demands a given activity makes on working memory dictate which encoding method, that is retrieval structure or elaborated memory structure or a combination of the two, as illustrated in Figure 4, individuals are likely to choose so as to attain reliable and rapid storage and access of information in LT-WM, and thereby determine the structure of the acquired memory skill.

The mechanism of STM accounts for working memory in unfamiliar activities, but does not appear to provide sufficient storage capacity for working memory in skilled complex activities. One possible explanation is that the general storage capacity is greater for specific skilled activities than unfamiliar ones. A more parsimonious account, however, is that storage in working memory can be increased and is one of many skills attained during the acquisition of skilled performance. Our proposal for LT-WM is consistent with this account. Also consistent with the conclusion that the superiority of expert memory and of exceptional memory is domain-specific, our proposal asserts that increased working memory is limited to the specific skilled activity in question. The acquired nature of LT-WM implies differences between tasks and in addition potential individual differences in the implementation of LT-WM for a given task. This raises new theoretical and methodological challenges for research on working memory to which we will now turn.

Toward a Theoretical Framework of Working Memory

According to our proposal LT-WM is not a generalizable capacity that, once acquired, can supplement ST-WM in any cognitive activity. LT-WM is acquired in particular domains to meet specific demands on storage and retrieval that a given activity imposes. LT-WM must therefore be discussed in the context of specific skilled activities. To provide this context, we first discuss the general structure of cognitive processes.

Cognitive processes can be described as a sequence of states or thoughts. Memory mediates between the states of this sequence. Cognitive states are dependent on each other, and memory generates this dependency, as do environmental correlations.

Memory plays another role that must be noted and differentiated. Thoughts - the cognitive states - are themselves the end products of complex generation processes. Typically, sensory and perceptual as well as conceptual operations are involved in the genesis of cognitive states, which requires knowledge activation and elaboration processes at various levels. In order for a higher level process to use the output of a lower level process, that output must remain available for at least some minimal amount of time. This availability is achieved through process-specific memory buffers that contain for a limited amount of time the results of the intermediate processes that generate the end product, or the cognitive state. We distinguish these buffers for the storage of intermediate results from the role of memory as a mediator between successive cognitive states.

Figure 5 illustrates our conception of the dual role of memory. For each cognitive state, there are complex generation processes at various levels of analysis, ranging from the sensory to the perceptual to the conceptual. The end products of these processes are the cognitive states that succeed each other over time: the varying contents of STM, the changing focus of attention, the flow of conscious experience. The arrow in Figure 5, which points to these end states, represents the complex, not necessarily linear sequence of processes involved in the generation of a cognitive state, including the necessary memory buffers for the temporary storage of intermediate processing results. These processes depend on the state of the environment as well as on long-term memory factors, namely the individual's experience and knowledge. Conversely, once a cognitive state has been generated, actions may occur that affect the environment, and traces of the present cognitive state may be retained in LTM, as indicated in Figure 5. Thus, to fully explain the succession of cognitive states, one would need to know (1) the state of the environment and its effects on the individual, (2) the individual's knowledge, experience, and beliefs, and how they interact with environmental effects, and (3) the previous cognitive state. According to this hypothesis, one would not need to know how this previous cognitive state was generated or the results of intermediate analyses. Thus, we assume that for the study of cognition the impact of neural activity can be summarized by a limited number of generated results or products.

Insert Figure 5 about here

This view of memory is not without precedents. Ever since Aristotle, complex cognitive activities such as thinking have been described as a sequence of thoughts. More specifically, Newell and Simon (1972) proposed that the contents of STM were sufficient to characterize the sequence of mental states corresponding to cognitive processes. Recently, Anderson (1987a) argued for the distinction of macro and micro processes, macro processes roughly corresponding to those processes that generate stable products affecting the sequence of the thought process. In his model of text comprehension, Kintsch (1988) proposed that successful text comprehension can be described as a sequence of states mediated by cycles of construction and integration of segments of the text.

Although the characterization of cognitive processes as a sequence of states is generally accepted, there is much more controversy over how the relevant information and intermediate products of a state are kept accessible. The modal view of working memory is that all of the relevant information is kept temporarily activated. Our account of LT-WM proposes that in skilled activities a significant part of the accessible information is stored in LTM and is accessible through retrieval cues in STM. It is very difficult to discriminate between these two accounts by mere observation of normal skilled processing. The two accounts could be differentiated, however, if the cognitive processes were interrupted at a given state and the subjects' attention were suddenly diverted toward another demanding, unrelated activity for some time until the former cognitive activity could be resumed again. If the interruption had a sufficiently long duration, the activated information in ST-WM would be irretrievably lost and the interrupted activity could not be continued. Information in LT-WM, on the other hand, would only become inaccessible and could be subsequently retrieved. If the significant information were stored in LT-WM, reactivation of the associated retrieval cues in STM would allow subjects to resume their cognitive process after the interruption. Hence designed interruption is an effective experimental technique to differentiate between storage in ST-WM and storage in LT-WM.

Successful experimental intervention in and disruption of cognitive processes require that the sequence of states in the normal cognitive process be well known and predictable as well as easily monitored. This is not the case for most cognitive processes studied in the laboratory. Typically subjects are given a task requiring the production of an answer, but the sequences of states for achieving this goal can vary greatly among subjects. Furthermore, it is difficult to determine the intermediate states in a given subject's cognitive process, which makes systematic intervention in and disruption of the process difficult if not impossible. There are, however, other types of tasks for which the sequence of intermediate states is better understood and controllable by external factors.

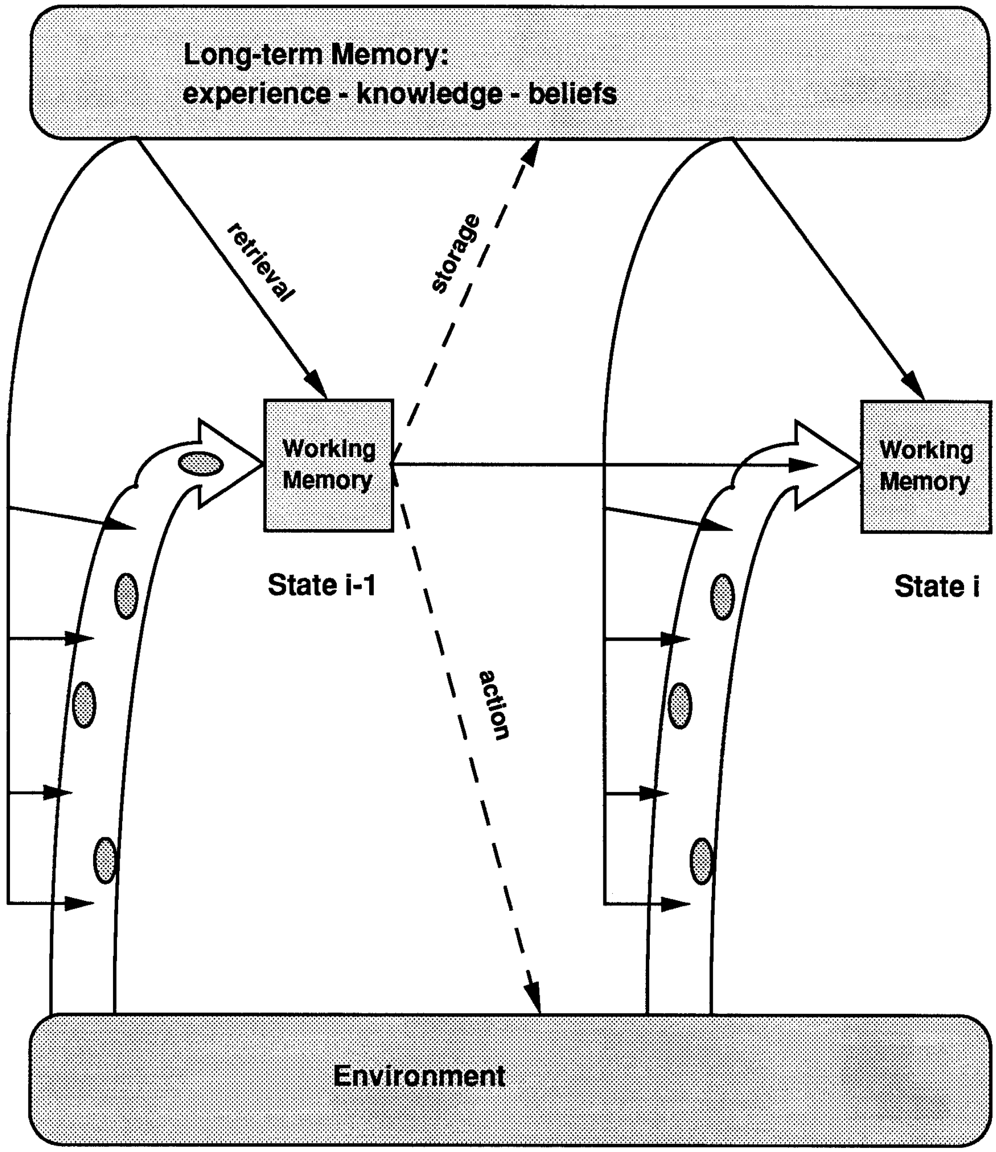
The task of comprehending a text is particularly well suited to the analysis proposed here. While reading a well-written text on a familiar topic, all skilled subjects read the text in more or less the same smooth and linear fashion. The sentences and phrases constitute natural boundaries of segments for processing. Even the words within a phrase are scanned in a linear, orderly fashion, a phenomenon which allows the study of the cognitive processes operating within segments. For all individuals who read, successful comprehension of a text involves the same predictable integration of information for sentences. Hence, as a first approximation, we can argue that comprehension of a text involves roughly the same sequence of states and segments for all skilled readers. In direct contrast, cognitive processes in most other skilled activities are constrained only by the act of successfully completing the task, and the intermediate steps and states may differ considerably across individuals.

Text Comprehension

Most educated adults have acquired high levels of skill in the reading and comprehension of texts. Hence text comprehension is far more common skill than the types of expert performance discussed earlier. Any findings about the structure of working memory during text comprehension could therefore be generalized more directly to the general population.

The goal of reading a text is typically to understand or comprehend it. When subjects have finished reading it is possible to assess successful comprehension by asking them questions about the text or by having them summarize the information presented in the text. Once they have understood a

Figure 5


Long-term Memory:
experience - knowledge - beliefs
retrieval
storage
Working Memory
State i-1
action
Working Memory
State i
Environment

test, subjects have substantial memory of the important aspects of the text even when the retention tests are delayed, a finding that indicates storage in LTM.

In the following review of text comprehension, first we examine the representation of text individuals construct and store in LTM. We then discuss the on-line processes during reading that produce this representation, describing the processes by which individual text segments are encoded and then explaining how the constructed representation of the previously read text is kept accessible in working memory so that the information can be encoded and successfully integrated with information presented earlier. We conclude the review by discussing the construction-integration model of text comprehension in order to provide a detailed example of the kind of retrieval structures we claim are the basis of LT-WM.

Memory after Completed Text Comprehension

There are several reasons for evaluating hypotheses about LT-WM by analyzing the memory individual s have for text after they have finished reading and comprehending it. Individuals' substantial memory for text reveals that storage in LTM during text comprehension not only is possible but also occurs as part of normal comprehension. A somewhat stronger connection between text comprehension and subsequent memory for ideas in the text is suggested by experimental manipulations of the degree of perceived comprehension, which is closely related to subsequent recall (Bransford & Johnson, 1973). In agreement with other investigators such as Gernsbacher (1990), we propose that a memory representation of the text is constructed during reading and stored in LTM. Furthermore we assume that the relevant parts of this representation remain accessible through cues in STM and thus form the basis for LT-WM during text comprehension.

The psycholinguistic literature has widely used a distinction, derived from linguistics, between three aspects of language: syntax, semantics, and pragmatics. Comprehension of a sentence is thought to start with a syntactic analysis, followed by a semantic interpretation, and possibly some pragmatic elaborations. According to this view, the end product of reading would be characterized at the syntactic, semantic, and pragmatic level. It has not been possible, however, to cleanly separate these processes in comprehension. We therefore doubt that this is a useful approach for analyzing the psychological process of comprehension, although we do not question its use in linguistic analysis. We prefer another distinction that is based more directly on the nature of the processes involved in discourse comprehension (Hörmann, 1976; van Dijk & Kintsch, 1983).

Different mental representations result from the linguistic processing of the language of a text, the understanding of the text as a separate mental object, and the integration of the textual information and background knowledge. It is useful to distinguish three different levels of representation:

1. Linguistic surface structure. This component of the memory representation of a text comprises the traces of the words of a text, not only by themselves but syntactically, semantically, and pragmatically interpreted in the context of a sentence.

2. Propositional textbase. The textbase is a coherent conceptual representation of a text and its structure. It contains both a microstructure and a macrostructure. Micropropositions are usually directly derived from the text (they correspond to phrases and sentences), or they may be the result of inferential processes (e.g., to bridge a gap in the coherence of a text). The macrostructure results from selection and generalization processes operating on the micropropositions. Macropropositions are thus partly cued directly by the text, and partly inferred.

3. Situation model. The model of the situation described by the text integrates textual information and background knowledge. This representation is not strictly propositional; it may be a spatial model such as a map. The situation model provides the reader with a basis for inference, elaboration, and action. In most cases, because the reader's purpose is to construct a new situation model or to update an already existing one, this model is a major component of the memory trace.

Researchers concerned with memory for text have focused on different levels of representation. In general, parsimony has dictated that accounts based only on the surface structure be simpler than accounts requiring memory for the propositional textbase and situational models. Glanzer, Dorfman, and Kaplan (1981) claimed that it was sufficient to consider the surface level, at least as far as STM was concerned. Most researchers were unable to accept such a restriction and differentiated at least between a surface level and some sort of semantic representation (Frederiksen, 1975; Kintsch, 1974; Meyer, 1975; Schank, 1972; Schank & Abelson, 1977). On the other hand, it has been repeatedly demonstrated that although surface representations are unavailable in many cases, subjects manage to solve experimental tasks very well on the basis of their situation model (Bransford, Barclay, & Franks, 1972; Bransford & Franks, 1971; Johnson-Laird, 1983).

The division of the representation of text into these three levels is based on the systematic differences in the strength of memory that have been observed among these levels of representation (Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986). If subjects read a text and are later tested with sentences from that text as well as with distracter items, their ability to differentiate between old sentences and meaning-preserving paraphrases provides evidence of surface memory. To the extent that subjects' answers contain more false negatives for a recognition test on paraphrases of sentences which actually appeared in the text than on inferences, subjects can be said to have retained the propositional representation of the text. Finally, an increase in the false negative rate for inferences compared with the rate for contextually related but non-inferable distracter sentences can be considered evidence for a representation at the level of the situational model. Under the conditions of the experiment by Schmalhofer and Glavanov (1986), for instance, the estimated strength of the surface, textbase, and situation model traces were d' = .14, .68, and 1.96, respectively. Subjects in this experiment were reading excerpts from a textbook on the computer language LISP. The results of this experiment provide clear evidence for the generation of a situational model, as memory for the surface form and the textbase is not sufficiently accurate to account for the observed retention. It must be emphasized, however, that the level of representation subjects prefer strongly depends on the nature of the text they read as well as on their reading goals.

Although these studies show the usefulness of distinguishing among various levels of representation in text comprehension, they give no justification for reifying these levels. For one study the surface properties of a text may be most interesting; for another, the textbase or the situation model. However, there is no such thing as a surface buffer, nor are there three separate levels of cognition. These are only analytic categories, perspectives we can bring to bear in analyzing texts. In one case we want to emphasize the surface relations among the text units; in another, the semantic or situation model relations. Furthermore, these three levels are not exhaustive. For example, in the study of algebra word problems, it is important to distinguish the algebraic problem model from the naive understanding of a situation (Nathan, Kintsch, & Young, 1992); in human-computer interaction, the distinction between a system model (a task analysis from the standpoint of the system) and the situation model (how the task is understood by the naive user) has proved to be useful (G. Fischer, Henninger, & Redmiles, 1991); in analyzing poetic language, the emphasis may be on rhyme, rhythm, and alliterative relations (Kintsch, in press-a). Thus, the levels of representation in text comprehension discussed here are to be understood simply as useful categories for the analysis of comprehension processes.

There are usually differences in the retention of the surface, textbase, and situation model of a text. The surface structure is generally stored until the end of a sentence and may be lost rapidly thereafter. The textbase is stored for the sentence currently being read and can be retrieved thereafter from LTM by means of conceptual retrieval cues. The situation model is often the longest lasting component of the memory trace (e.g., Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Although it is generally true that meaning is retained better than surface memory (Bransford & Franks, 1971; Sachs, 1967; for sign language, Hanson & Bellugi, 1982) long-term retention of surface form is by no means rare (Hjelmquist, 1984; Kintsch & Bates, 1977; Masson, 1984). Indeed, surface form is retained best when the way something is expressed is pragmatically significant and thus relevant to the situation model. It matters a great deal whether a partner in a discourse has said something politely or aggressively, and in these situations the wording is quite well remembered (Bates, Kintsch, & Fletcher, 1980; Keenan, MacWhinney, & Mayhew, 1977). Outside of a social context, however, in laboratory studies of memory for sentences, memory is in general propositional, and surface features are typically reconstructed (Potter & Lombardi, 1990).

Other types of evidence also imply that skilled readers construct an integrated representation of the text in LTM. After reading a text, subjects are able to answer comprehension questions even when the answers require integration of information presented in different parts of the text (Daneman & Carpenter, 1980). Subjects are also able to extract the central message of the text and to provide valid summaries of the texts (Kintsch & Kozminsky, 1977). When instructed to recall a text, subjects are more likely to recall central ideas and propositions that are causally connected to many other propositions in the text (see van den Broek [1990] for a review) and that are at higher levels in the hierarchical structure of the textbase (Kintsch & van Dijk, 1978).

In a recent book Gernsbacher (1990) has assembled a wide range of evidence supporting the notion of structure building. She shows that the pattern of subjects' reading times is consistent with building a structure of text because more reading time is required to lay the foundation for a new structure or sub-structure than to continue processing within the current structure. The constructed structure appears to be quite general, since it makes surprisingly little difference whether a story is presented as a regular text or as a sequence of textless cartoons (Baggett, 1975) or as a silent film (Baggett, 1979). Furthermore, text comprehension during reading is highly related to comprehension during listening (see Gernsbacher, 1990, for a review). This invariance over the type and medium of presentation suggest that the essence of comprehension involves situation models.

In sum, the evidence reviewed supports the claim that a structure of the text is constructed in LTM during reading. For this structure to be continually expanded to integrate new information from the text, relevant parts of it must remain accessible during reading. Our main conjecture is that the accessible portions of this structure in LTM serve as an extended working memory (LT-WM). Unlike some related proposals (cf. Gernsbacher, 1990) that assume that accessibility of all information in working memory is based on short-term activation, our proposal distinguishes the short-term activation involved in interpreting text segments from the storage and integration of final encodings in the structure of the text in LTM.

Memory During the Reading of a Text

When subjects read a well-written text, they proceed smoothly from one sentence to the next. We assume that when the processing of an entire clause or sentence is completed, the new information has been successfully integrated into the structure of the previously read text in LTM. When the next sentence is processed, some elements of the current structure of the text are kept in STM to provide context as well as to serve as retrieval cues for the accessible portions of that structure. We make an important distinction in our proposal between the final state corresponding to a completed encoding of a new text segment and the transitory period when the segment is being read and processed. We now consider memory during reading and describe our proposal in more detail. We first discuss the short-term storage of a text segment while it is being processed. Then we consider how elements in the structure of a text can serve as context that allows information to be integrated across the boundaries of text segments.

Intermediate Representation of Text Segments

The sequence of intermediate representations generated during reading and the associated storage buffers have been explored in great detail (see, for instance, Potter, 1983). First, light receptors register visual information on the retina; then, in a series of transformations, the neural information is further analyzed en route to the associative cortex. Higher levels of representation are derived from the lower level representations as the information at each lower level is maintained briefly in associated memory buffers. One may thus describe the role of memory in the generation of cognitive states, as shown in Figure 5, as a sequence of temporary memory buffers with different characteristics. What is stored in these memory buffers are temporary representations, which are accessible to other cognitive processes only in highly restricted ways. Some low-order buffers can be accessed only by certain specialized processes. Their contents are never consciously experienced, and their existence can be demonstrated only by special experimental methods. The contents of other buffers, on the other hand, may be transferred to the level where conscious experience is possible; but at that point they become contents of working memory, characterized by the same properties as other contents of working memory. Thus, some results of intermediate perceptual and conceptual processing are only temporarily buffered and by their very nature are transitory and unconscious. Others are normally so, but may become cognitive end products in their own right and are thus consciously experienced.

Furthermore, some of these intermediate buffers receive input not only from the sensory, perceptual, and conceptual processes but also from working memory directly. That is, they serve as slave systems for the central executive, as Baddeley (1986) has shown. These are the articulatory loop and the visual-spatial scratch pad, which are capable of recycling special sorts of information under the control of the central executive.

Thus, putting some of the notions of Potter (1983) and Baddeley (1986) together, we arrive at the following sequence of intermediate buffers in reading a text. The neurophysiological bases of the first three buffers including the differentiation between space and pattern information are relatively well understood. The remaining memory buffers with higher levels of representation involve learned recodings of visual stimuli.

1. Retinotopic icon. This buffer is the perceptible continuation of a single presented word resulting from photoreceptors and other neural mechanisms. Apparently, this buffer has no function in normal reading and plays a role only under special laboratory conditions.

2. Spatiotopic visual memory. Retinal information from successive eye-fixations is integrated at this level of processing. The printed text is represented as a stable structure located in space.

3. Reatopic visual memory. Spatial characteristics of the retinal information are less relevant at this level of transformation than are the configuration of visual features and patterns of the perceptually available text segment. Even when this text segment is removed and replaced with an irrelevant stimulus (visual mask), information about the original text segment is retained in this memory buffer for several seconds.

4. Conceptual buffer. At this level of representation, words and objects are perceived and understood. Experimental results show that words can be momentarily understood, but are then lost because of interference from other cognitive processes. Conceptual representations become conscious and receive permanence only when transferred to working memory.

5. The articulatory loop. Even skilled, adult readers transform visual information in reading into an acoustic form (we ignore here the debate whether this level of representation is best understood as acoustic or articulatory-motoric). Articulatory representations not only are intermediate results in the sequence of cognitive processes but also may become contents of working memory; and material from working memory can be recycled in the articulatory loop. Thus, this buffer may function as a slave system for working memory (Baddeley, 1986).

6. The visual-spatial scratch pad. This is Baddeley's (1986) other slave system, which has much the same function in the visual-spatial domain as the articulatory loop in the acoustic domain.

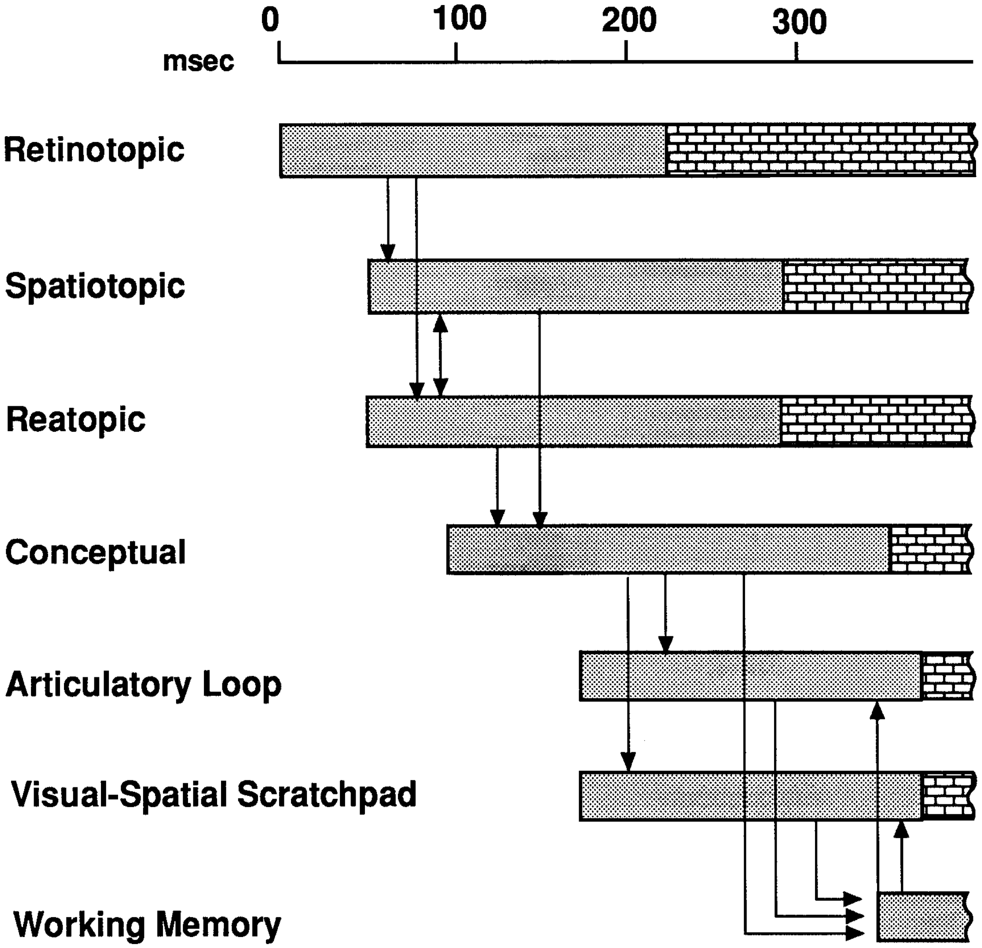
7. Working memory. This is comparable to Baddeley's (1986) central executive. It comprises the entire cognitive work space (here we deviate from Potter, 1983) where information about the current text segment is stored in rapidly accessible form. Information in working memory is the cognitive end product of all the foregoing processes, but is itself complex and structured and involves different levels of representation.

Figure 6 (modified from Potter, 1983) illustrates this temporal sequence of buffers. Positron emission tomography (PET) and event-related potential (ERP) studies indicate activity in the right posterior cortex during visual feature analysis (Posner & McCandliss, 1993), beginning at about 100 ms. By 200 ms there is activity in the left-lateralized visual word form area if the strings analyzed are word candidates, but not otherwise. The time necessary to encode acoustical information can be estimated as about 175 ms: A reader needs at least 400 ms to pronounce a word, 225 of which appear to be needed for articulation (if a subject knows which word is to be pronounced, the time to articulate it is 225 ms). Conceptual processing begins about 100 ms after the fixation of a word. At 250 ms, PET and ERP results show diffuse activation in frontal brain regions of both hemispheres. Priming studies (e.g., Swinney, 1979) indicate that it takes about 350 ms for the meaning of a word to be fixated in context. PET studies indicate that if subjects are given enough time, the activity associated with semantic processing shifts from the frontal regions to posterior regions (Posner & McCandliss, 1993).

Insert Figure 6 about here

The sequence of buffers outlined in Figure 6 is critical to the integration and comprehension of segments of text (cf. Figure 5), such as phrases and sentences. Deficits in any one of these storage buffers may have complex results beyond the buffer in question. Thus, phonological storage deficits affect not only the acoustic buffer itself but also general learning and comprehension capacities. Baddeley et al. (1988) observed a patient with such a deficit and found a normal ability to associate

Figure 6


0
100
200
300
msec
Retinotopic
Spatiotopic
Reatopic
Conceptual
Articulatory Loop
Visual-Spatial Scratchpad
Working Memory

meaningful, familiar word pairs, but an almost total inability to associate words with unfamiliar sounds (words in another language). Apparently, the patient could process familiar sound patterns adequately, without having to maintain them in the acoustic buffer. But the patient did not have enough time to analyze unfamiliar sound patterns without maintaining them at least briefly in the acoustic buffer. Baddeley and Wilson (1988) observed a similar patient who had no trouble identifying single words and comprehending brief sentences, even syntactically complex ones. This patient, however, could not understand longer sentences because, although comprehension abilities were intact, the patient could not maintain speech sounds long enough in the acoustic store to complete a successful analysis. Analysis was unsuccessful, for instance, when an early word in a sentence could not be disambiguated until late in the sentence.

Although Figure 6 represents our best guess at this time, we are not strongly committed to this particular sequence of buffers and their time course, or to the way the arrows between these buffers have been drawn. Further research will change some of this detail; the important point for us concerns the existence of some such sequence of intermediate buffers and their relation to working memory. It is, of course, a highly significant research question to come up with a more accurate version of Figure 6 - but it is not our question. We merely want to raise this issue and show how it relates to the study of working memory proper. We can now refer to what Baddeley has called the central executive without having provided definitive answers to the many questions about intermediate memory buffers, but having shown, at least, how this aspect of human memory fits in with our view of working memory.

The Time Course of the Construction of Mental Representation of Text Segments

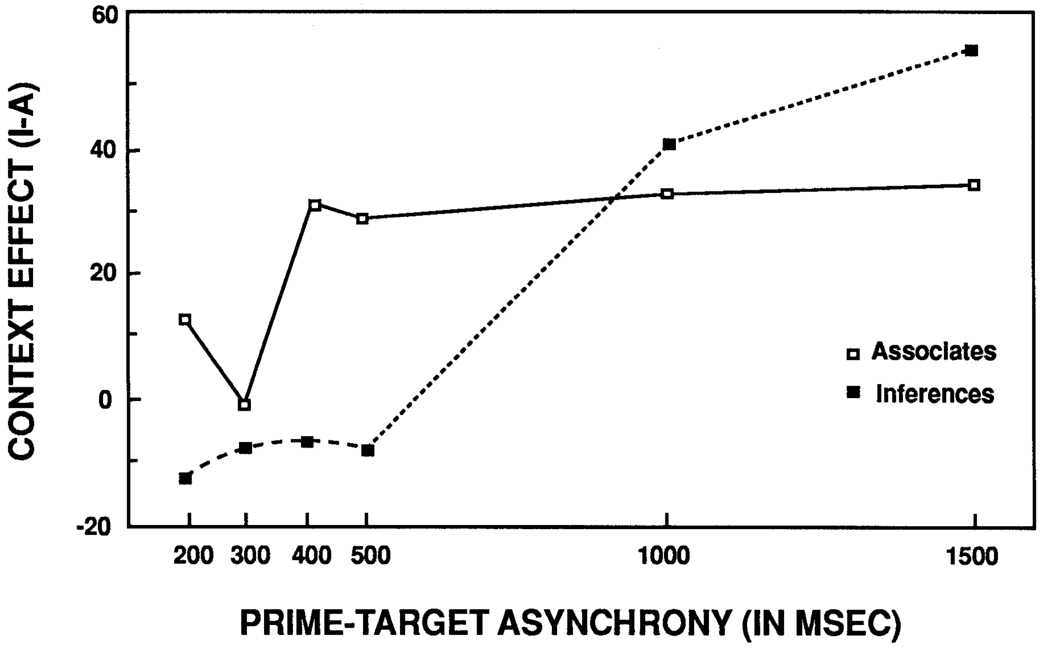
It is necessary not only to distinguish different levels of representation in the cognitive states that arise in reading but also to be aware that the time to generate these different end products may vary. As shown in Figure 6, it takes about 350 ms for a word to be made available in working memory. At this point the word has been acoustically coded and identified semantically. Other processes involved in text comprehension, such as the generation of a situation model, may take considerably longer, however. An experimental study by Till, Mross, and Kintsch (1988) illustrates this fact. This study explored the relationship between a homonymic priming word embedded in a disambiguating sentence and target words that are associates either to the context-relevant meaning of the homonym (e.g., the priming word was "mint" used in the sense of "building," and the target "money") or the context-inappropriate meaning (e.g., the target "candy"). The results are shown in Figure 7, in which the difference in the reaction time between unrelated and related target words is plotted as a function of the prime-target asynchrony. For the first 350 ms after the prime word was presented, reaction times to related and unrelated target words did not differ; but after that time, reaction times for related target words were consistently faster.

These data replicate the results of Swinney (1979), Gernsbacher, Varner and Faust (1990), and others. They imply that context does not work like a filter that facilitates the expected or inhibits the unexpected. Rather, the construction of meaning in a discourse takes time and requires the integration of the target word and its context.

Insert Figure 7 about here

It takes considerably more time, however, to construct a situation model than to disambiguate a word. For the simple sentences used in the Till et al. study, the construction of a situation model amounted essentially to inferring the sentence topic. Apparently, this inference was not completed until 1 s after the end of the sentence. Given the sentence "The townspeople were surprised that all buildings had collapsed except the mint," subjects could not identify "earthquake", which was in the sentence topic as determined by a prior rating study, as a word more rapidly than they could identify thematically irrelevant control words when the words were presented 200-500 ms after the end of the sentence, even though it would have been possible for subjects to have inferred the sentence topic earlier while reading the sentence. As Figure 7 shows, only after 1 s could a priming effect be observed. From this result we conclude that 1 s was needed to infer the sentence topic for isolated sentences. Word meanings are formed quickly; discourse meaning, in this case, took about 1 s to be constructed.

Figure 7



60
40
20
0
-20
CONTEXT EFFECT (I-A)
200
300
400
500
1000
1500
PRIME-TARGET ASYNCHRONY (IN MSEC)
□ Associates
■ Inferences

A similar emergence over time for the encoded structure of sentences has been documented by Gernsbacher (1990). When two participants are mentioned in a sentence, there is an immediate access advantage for the most recently mentioned participant, but over time (1-2 s later) the advantage is reversed and the first-mentioned participant is more accessible. According to Gernsbacher's (1990; also Kintsch & Welsch, 1991) conception of text comprehension as structure building, the first-mentioned participant is part of the structural foundation to which subsequently mentioned information is related. Subsequent retrieval, once the transient activation has subsided, is mediated by the constructed structure, which gives the first-mentioned participant an access advantage.

Converging evidence for the time course we propose for construction of mental representations in reading comes from studies with PET and ERP methods that are reviewed by Posner & McCandliss (1993). These authors point out that visual feature and word-form processing appear to be entirely bottom-up and occur without previous activation in phonological or semantic areas. At about 250 ms, there is very diffuse activity in frontal brain regions critical for semantic processing. Presumably, this activity corresponds to the initial activation of various possible semantic codes associated with the visual word form that has been determined by that time. Later, semantic activity shifts to posterior areas (Wernicke's area) and becomes hemisphere specific: Whereas the diffuse activity persists in the right hemisphere, all but the most common associates are suppressed in the left hemisphere. Presumably, this ongoing activity corresponds to the contextual fixation of word meanings - that is, the suppression of context-irrelevant information - and the construction of an elaborated discourse meaning, such as the formation of topical inferences.

In general, readers attempt to interpret whatever structure they encounter - lexical, syntactic, semantic, or discourse - immediately rather than engaging in a wait-and-see strategy (for a detailed review, see Just & Carpenter, 1987.) However, this means only that they begin the interpretive process as soon as possible. It is not necessarily the case that each word is fully encoded while it is being fixated. A fixation lasts around 200 - 250 ms (See Footnote 1), whereas word meanings in a discourse context need around 350 ms to stabilize. It has long been recognized in the case of syntactic analysis and text level processing that while these processes start immediately, they may take a considerable amount of time, in part because the required information is provided only at a later time in a sentence or a discourse (Aaronson & Scarborough, 1976; Baddeley & Wilson, 1988; Just & Carpenter, 1987). The same is true, however, for the encoding of word meanings, which also may take longer than the time it takes to fixate the word (even though the eye movement data reported by Just and Carpenter provide no support for this claim).

We have argued that it is necessary to differentiate the function of memory in the process of generating cognitive states from its function in relating different states. In the former case, memory buffers contain intermediate results, which are significant for the formation of the cognitive state, but irrelevant once it has been formed. In the other case we are talking about the storage and retrieval of cognitive end products. Focusing on reading, we have specified the buffer sequence that is necessary for the retention of intermediate results. Further, we have emphasized that the cognitive state that is the end product of reading, is complex and must be characterized at different levels of representation, and that different time parameters are associated with the generation of these different levels. These lengthy arguments were necessary because if we want to talk about retrieval and storage of cognitive end products in a complex process such as reading and to differentiate these processes from the buffering of intermediate results, we must clearly and adequately characterize the cognitive states that are achieved in reading. For instance, the findings that readers remember the surface features of a text does not necessarily mean that they remember intermediate results. Surface features can be a component of a final cognitive state just as well as other aspects of the text.

Working Memory Across Boundaries of Text Segments

We now consider how the end products encoded in the structure of the text can remain accessible while subsequent segments of the text are processed. The central characteristic of text comprehension is the integration of successive sentences of a text into a coherent representation. Information about text previously read must be kept accessible in working memory if the sentence currently being read is to be successfully integrated. Numerous questions can be posed about the amount and type of information that must be maintained in an activated state in STM-WM and kept accessible in LTM-WM during text comprehension. Here a formidable methodological problem arises.

According to the prevailing view, working memory in text comprehension is based only on transient activation of information. The prediction following from this view is that a disruption of reading and engagement in an unrelated, attention-demanding activity should lead to an irretrievable loss of the information in ST-WM necessary for continued text comprehension. When reading is resumed at a later point, comprehension of the following text should be greatly impaired. It is virtually impossible, however, to study the accessibility of information in working memory during text comprehension without some kind of additional testing or other type of disruption of the normal continuous reading process.

If, on the other hand our proposal for LT-WM is correct, this methodological problem dissolves. The prediction that follows from our proposal is the opposite of the art that follows from the prevailing view: Disruption of text comprehension should not cause irreparable harm and can be corrected by retrieval of the necessary structures from LTM and by reactivation of the corresponding retrieval cues in STM. If so, a broader range of techniques can be used to study working memory. In this section, after reviewing disruptions, we consider studies that have examined access and activation of information in working memory.

The effects of disrupted reading on text comprehension. In a long series of experiments Glanzer and his colleagues (B. Fischer & Glanzer, 1986; Glanzeret al., 1981; Glanzer, B. Fischer, & Dorfman, 1984) have interrupted sentence-by-sentence reading of texts with an unrelated activity and then allowed reading of the text to resume. When Glanzer and his colleagues compared interrupted reading to a normal control condition, they found a very consistent general pattern of results across different experiments with different types of interrupting activities. Most important, disruptions of text comprehension did not reliably influence either the speed or the accuracy of answers to comprehension questions. The effect of the disruption was limited to an increase in the reading time for the first sentence after reading was resumed. These two findings imply that information in working memory was not irreparably lost during the interruption. By retrieval of the appropriate elements in the structure of the text from LTM, it should be possible to reinstate the context and the accessibility of information relevant to continued comprehension.

The ease and speed with which critical elements are retrieved after the interruption should be influenced by the interpolated activity. Because the increased reading time after the interruption should reflect these retrieval times, we examine the increased reading times as a function of the type and duration of the unrelated activity during the interruption.

In the first experiment Glanzer et al. (1981, Exp. 2a) compared continuous reading to an interruption involving the solution of three short addition problems and found an increase in the reading time of 450 ms. The reading time for the three-problem intervention was longer by 250 ms that for the one-problem intervention. Although the duration of the interruptions was not recorded, it can be estimated to be less than 10 s even for the three-problem intervention. In later studies the duration of the interpolated activity was extended to around 30 s. B. Fischer and Glanzer (1986, Exp. 2) used two trials of a digit recall task, which lasted an average of 31 s, as the interpolated activity and with this activity B. Fischer and Glanzer obtained an increased reading time of around 1200 ms for resumed reading. In two other experiments (Exp. 3 and 4) they had subjects verify additions for around 30 s and obtained an increased reading time of around 1800 ms. Both the duration and particular type of intervening activity appear to influence the accessibility of the information about previously read text.

Glanzer also studied the effect of intervening comprehension of unrelated sentences and different texts, material which would be expected to be maximally interfering. In one experiment (Glanzer et al., 1984, Exp. 1) sentence-by-sentence reading was interrupted by the tasks of reading sentences about unrelated facts, which took around 7 s. The resumed reading time increased by 314 ms. The reading time for the intervening unrelated sentences and subsequent memory for these facts were no different from those observed for a control condition. A subsequent experiment (Glanzer et al., 1984, Exp. 3) increased the duration of the unrelated text comprehension task to around 30 s by having the subjects alternate between two texts. The increase in the time subjects took to resume reading of the primary text was 355 ms. When the unrelated activity was comprehension of a different text, accessibility of information in working memory appeared to remain high even when the duration of the unrelated activity was increased.

In summary, the research by Glanzer and his colleagues shows that the transient portion of working memory (ST-WM) is not necessary for continued comprehension of the type of texts they studied. Their findings are consistent with the view that the necessary information is stored in LTM-WM, where interruptions lead to a loss of the necessary retrieval cues in STM. When the intervening task involves comprehension of unrelated sentences or texts, the additional time for accessing the retrieval cues that allow continued reading is around 350 ms, which is roughly equivalent to other estimates of retrieval from LTM reviewed earlier. When the intervening task is unrelated to reading and has a comparatively long duration (around 30 s), access times are longer and may involve reinstating the task of reading (B. Fischer & Glanzer, 1986; Glanzer & Nolan, 1986).

The hypothesized role of storage of surface structure in STM. The finding that subjects' text comprehension does not seem to be impaired by long interruptions between the reading of consecutive sentences in the text seems to rule out any significant role of transient information in ST-WM. However, this observation runs counter to a long tradition in text comprehension and discourse analysis. Sentences in a text contain references, such as pronouns, to previously presented information and cannot be fully comprehended unless they are successfully resolved. Some researchers have hypothesized that access to the surface form of preceding sentences in ST-WM is necessary for successful linkage between the current sentence and preceding sentences. Glanzer and his colleagues (Glanzer et al., 1981; Glanzer et al., 1984) clearly formulated this idea in their early work. They proposed that what is carried over in short-term memory is the linguistic surface form of the preceding sentence or sentences, uninterpreted semantically or pragmatically.

Investigators have tried to determine whether intermediate linguistic structures are retained between sentences rather than, or perhaps in addition to, the sentence meaning by examining how much of the preceding sentences subjects can recall verbatim. If subjects are interrupted during reading at randomly selected places and asked to reproduce what they have just read verbatim, they reproduce the last sentence or clause almost perfectly, and the next to the last clause fairly accurately, but cannot produce earlier sentences and clauses (see Jarvella, 1979, for a review). This replicable finding proves only that readers can reproduce most of the surface form of two clauses. Because the subjects in these experiments knew that they would be tested for short-term retention, it is very likely that they used special chunking and rehearsal strategies. Two sentences is therefore almost surely an overestimation of the contents of what is retained verbatim during normal reading, just as the immediate memory span is an overestimation of short-term capacity during list learning (Glanzer & Razel, 1974) or reading (Daneman & Carpenter, 1980). In fact, when subjects are informed in advance of the way they will later be tested, their speed and pattern of reading are dramatically influenced by the particular test they anticipate (for example, Aaronson & Ferres, 1984; Kieras, 1984). Thus, the verbatim recall data do not necessitate the assumption that readers hold in memory the uninterpreted surface form of a couple of sentences.

In an effort to demonstrate empirically the role of the surface form of the two preceding sentences in ST-WM, Glanzer et al. (1984) explored the conditions under which the increased reading time after disruption could be eliminated. If, following the interruption, subjects reread the last two sentences before proceeding to the next sentences, no reliable increase of the reading time was observed in Exp. 3. Experiment 4 showed that rereading only the last sentence was sufficient to eliminate the increase in reading time associated with the interruption. In a post hoc analysis of their texts, Glanzer et al. (1984) found that many of the sentences subjects first encountered after the interruption did not have references to the preceding sentences (independent) and therefore did not require linkage information in ST-WM. A reanalysis suggested that the increase in reading time after interruption was greater for dependent sentences which would require linkage information, than for independent sentences. For independent sentences, rereading another independent sentence from the previously read text--not necessarily the last sentence read before the interruption--was found sufficient in Experiment 5 to eliminate the increase in reading time after the interruption. In a subsequent study B. Fischer and Glanzer (1986) systematically varied the dependence and independence of the sentences in their texts and were able to demonstrate larger increases in resumed reading time for dependent than for independent sentences. From the results from their four experiments, B. Fischer and Glanzer (1986) estimated additive increases in resumed reading times to be 408 ms when the increase was due to unavailability of general theme and to be 402 ms when the increase was due to unavailable linkage information (the latter finding only to dependent sentences).

According to our proposal for LT-WM, these results should be interpreted as reflecting the reinstatement of the retrieval cues necessary for accessing the hierarchical structure readers have generated for the previously read text in LTM. Rereading the last sentence of a text is sufficient to activate retrieval cues to the general structure as well as cues pointing to specific information corresponding to the last sentence. Rereading any independent sentence of the prior text provides access only to the general structure (cf. Glanzer et al., 1984, Exp. 5). When there is no opportunity to reread prior sentences, access to the general structure requires around 400 ms; and for dependent sentences, retrieval of the specific information prior to the interruption requires a similar amount of time.

The interruption procedure, therefore, does little more than slow down the reading process a bit, because what is disrupted is neither a perceptual trace nor purely a surface representation of the text, but a fully analyzed, fully interpreted representation of the previous text in long-term memory that is readily accessible in long-term working memory.

ST-WM during reading. Text comprehension has been shown to involve the generation of an integrated representation in LTM of the previously read text. Access to these structures in LTM is mediated by retrieval cues transiently activated in ST-WM. These findings are consistent with our proposal for LT-WM. In this section we argue that storage of these retrieval cues is also consistent with the capacity limits of ST-WM. We review research on the effects of other concurrent activities, some of them explicitly designed to interfere with ST-WM. We then discuss some research directed toward measuring the accessibility of information in working memory.

There are several ways in which retrieval cues to the integrated representation of the text may be coded and maintained in ST-WM, which consists of several types of buffers. The most direct approach to determining how these retrieval cues are stored is to selectively interfere with a buffer by forcing subjects to perform an additional task concurrently with text comprehension. To interfere with the articulatory loop, researchers can instruct subjects to vocalize some unrelated verbalization during reading. This manipulation does not seem to impair normal text comprehension, and decreased comprehension is observed only with difficult text for which information about word order has to be preserved. In a review Baddeley (1986) concluded that the articulatory loop is not typically needed for text comprehension by skilled readers and is used as a supplement to aid in comprehension of difficult text. Additional support for the robustness of comprehension comes from a frequently replicated finding that subjects can read a text aloud without impairing their memory or comprehension of the text (for a review see Ericsson, 1988b), although reading aloud is somewhat slower than silent reading, especially for skilled readers.

The clearest evidence for the distinction between traditional STM and the transient storage of information during text comprehension comes from studies by Baddeley and his colleagues. Subjects performed a STM task concurrently with a task requiring comprehension of sentences and texts. Even when considerable additional information was maintained in STM, comprehension performance was sustained with only minor decrements.

When subjects are asked to hold six digits in memory while reading a sentence and later to reproduce those digits, Baddeley and Hitch (1974) found a decrement in performance of about 10% compared to the performance of subjects who do not have a memory load. In a second experiment, subjects had to memorize three or six visually presented digits while listening to a prose text. With only three digits to learn, subjects answered questions about the text almost as well as a control group did (4% less). With six digits, a significant performance decrement of 18% was obtained (Baddeley & Hitch, 1974). Performance decrements exist for comprehension as well as for memory (Baddeley, 1986). In one of Baddeley's studies, subjects had to verify visually presented sentences (as in semantic memory experiments) while remembering from zero to eight spoken digits. The frequency of error increased only when subjects had to remember six or more digits, whereas reaction times increased modestly with memory load.

Results like these show that in dual task situations, text comprehension is relatively unaffected by low and intermediate loads on memory and only really impaired when the resource demands of the secondary task are maximal: Even without anything else to do, individuals can manage to remember six or eight digits at most. As long as they have some free resources, however, they perform remarkably well on text comprehension tasks. These findings imply that the storage of information that makes the representation of the prior text accessible during text comprehension needs to reside in the central executive, according to Baddeley's (1986) model.

In addition to previously discussed findings on activated information during reading through priming several investigators have asked subjects to think aloud during reading. In a review of think-aloud studies Ericsson (1988b) found no evidence that standard think-aloud instructions (Ericsson & Simon, 1993) influenced comprehension compared to silent control subjects. At the same time subjects thinking aloud while reading tend to give very sparse and uninformative reports for well-written easy texts and subjects predominantly read the text aloud. This pattern of verbalizing primarily the presented text during fluent comprehension and only in the case of comprehension difficulties verbalizing deliberate efforts to attain understanding is consistent with our framework (Kintsch, 1988). To gain more information about the stable end-products of comprehension researchers have modified the reading task in think-aloud studies and stop the subjects after each sentence to comment and verbalize the information associated with that state. These verbal reports reflect access to information stored in LT-WM, which we will turn to next.

<u>LT-WM during reading.</u> Evidence for accessibility of the representation of the prior text is largely indirect because successful comprehension would be impossible without such access. A few studies have measured accessibility during text comprehension in a relatively direct fashion, however. While reading subjects been interrupted between consecutive sentences and have been presented specific probes for cued recall or recognition of previously read information, or asked to verbalize their thoughts (think-aloud) or directed to comment or elaborate on the text. Although such interruptions have typically been viewed as very disruptive and incompatible with normal text comprehension, the research we have previously reviewed here on more demanding and unrelated intervening activities showed no reliable evidence of effects on final comprehension.

A small number of studies have compared the effects of thinking aloud and directed commenting with silent reading. No evidence was found that mere thinking aloud influences comprehension and memory of the text, but additional directions to elaborate on the sentences in the text leads to better memory for the text (see Ericsson, 1988b, for a review). When subjects think aloud about a recently read sentence, they verbalize the information in attention (Ericsson & Simon, 1993), and further verbalizations should reflect the most directly accessible information. By analyzing the propositional information from the text contained in think-aloud verbalizations, Fletcher (1986) was able to evaluate theoretical models that explain how information is selectively retained in ST-WM to maximize the generation of a coherent text representation. Guindon (1980, 1981) focused her analysis of the think-aloud protocols on the access and generation of verbalized inferences that went beyond the information explicitly provided by the text. A recent series of studies by Trabasso and Suh (1993) represents the most comprehensive effort to analyze information verbalized after each sentence. Trabasso and Suh successfully predicted the frequencies with which specific inferences were verbalized. Most important, they were able to show that verbalized information from one group of subjects could successfully predict priming and retention for different groups reading the same stories silently.

During encoding and storage of the current sentence, the relevant information from the previously read text must remain accessible. Retrieval cues to the hierarchical organization of the encoded text provide access to this information, but direct access is limited to recent information as well as to elements central to the structure. A few studies provide empirical support for this prediction of differential accessibility. Fletcher (1981) tested Kintsch and van Dijk's (1978) formulation of this prediction by interrupting subjects during text comprehension and having them make recognition judgments for arguments of propositions from sentences they had previously read. Recognition times were about 200 ms faster for central propositions that according to Kintsch and van Dijk, should remain accessible, than for other propositions. The speed of access to the central propositions matched that of recently presented propositions, a result suggesting a similar state of immediate accessibility.

Using a related distinction between information describing the topic of a text and information describing the details, Nolan and Glanzer (reported in Glanzer and Nolan, 1986) conducted two experiments in which they interrupted subjects were interrupted during reading and had them make recognition judgments of presented sentences. In the first experiment, Nolan and Glanzer contrasted the currently presented information for both topic and details with information presented three sentences earlier. They found no difference in the recognition times for topic information, but recognition times for previously presented details were around 700 ms longer than for details in the current sentence. Hence, details from previous sentences are not kept directly accessible, but require mediated retrieval.

In a subsequent experiment, recognition of topic and detail statements was compared for information presented in the current paragraph and in the previous paragraph. The only reliable effect was that recognition of details in both paragraphs took around 500 ms longer than did recognition of topics. Accessibility of topic information in both experiments remained high.

Nolan and Glanzer (Glanzer & Nolan, 1986) found an intriguing effect on reading times for reading resumed after interruptions with recognition tests. When the recognition test involved topic information from the previous paragraph, the reading times were over a second longer than for topic information from the current paragraph. The same pattern of reading times occurred in recognition tests of details from the previous and current paragraph, although recognition tests of details were associated with 300 ms of additional reading time compared to tests of topic information. This finding clearly suggests that retrieval of prior topic information requires access to and reinstatement of the current structure so that comprehension can continue.

Individual differences in comprehension: encoding skill or general capacity. Several explanations have been proposed for the large individual differences that have been found in text comprehension. The most plausible loci for these individual differences are the capacity to maintain previously presented information in ST-WM or alternatively the acquired skill to encode earlier presented information in accessible form in LT-WM. We will first briefly discuss the traditional view based on differences in the capacity of ST-WM, which has been the major theoretical mechanism to motivate empirical research. Then we will reevaluate the empirical evidence to show that it supports our proposal for skilled encoding and retrieval of information from LT-WM. The evidence we have reviewed so far shows that during text comprehension readers generate relevant inferences and form an integrated representation that is highly resistant to short-term interruptions and thus does not involve ST-WM for storage.

Consistent with Baddeley and his colleagues' observations of only mild interference from a concurrent task during reading, several investigators have been unable to account for individual differences in text comprehension based on individual differences in performance on standard tests of the capacity of STM. Furthermore, dramatic increases in text comprehension from childhood to adulthood do not correspond to comparable increases in the capacity of STM (Case, 1978; Chi, 1976; Dempster, 1981; Huttenlocher & Burke, 1976); there are no reliable differences in memory span for good and poor readers (Farnham-Diggory & Gregg, 1975; Rizzo, 1939). These results are inconclusive, however, because the memory span is purely a test of storage capacities for unfamiliar and unrelated information and does not indicate the capacity of working memory available during reading. Working memory has a dual function, processing as well as maintenance. Hence, even an account based on LT-WM requires sufficient capacity of ST-WM to allow retrieval from and encoding of the presented text in LTM as well as maintenance of necessary retrieval cues. Problems with decoding, accessing the meanings of words, retrieval of relevant knowledge, and necessary inferences should all lead to difficulties in the smooth continuation of text comprehension.

Daneman and Carpenter (1980) therefore designed a task to measure the capacity of working memory during reading. They presented subjects with a series of unrelated sentences which they needed to comprehend in order to answer subsequent test questions. At the end of the presentation, subjects were asked to recall as many of the last words of the sentences as possible. The number of words a subject correctly recalled is the subject's reading span. College students manage to recall the last word from one to five sentences; good readers have a significantly higher reading span than poor readers. Reading span correlates with comprehension tests ($r = .5$ to $.6$) and with the ability to answer content questions about a previously read text ($r = .7$ to $.9$).

Some of the questions required the readers to successfully resolve the referents of pronouns in the text. Subjects with high reading spans gave very accurate answers regardless of the distance between the pronoun and its earlier referent in the text. Subjects with lower spans, however, were less accurate, and the frequency of errors increased as a function of the distance. The correlation between overall accuracy for resolving referents of pronouns and reading span ranged between 0.8 and 0.9. Similar correlations were obtained when the questions involved retrieval of facts presented in the text. Hence, ability to resolve referents of pronouns and ability to recall presented facts are both closely related to reading span. In retrospective reports and observations from the reading-span task, Daneman and Carpenter (1980) found evidence for the subjects' active efforts during reading to encode associations between the last words of the sentences as well as efforts "to reconstruct the sentences on the basis of whatever "gist" had been retained" (p. 457). If superior comprehension is viewed as efficient storage in LTM of information in the text, then the reading-span task may reflect shared skills and mechanisms.

More recent research by Engle and his colleagues (Cantor, Engle & Hamilton, 1991; La Pointe & Engle, 1990; Turner & Engle, 1989) have examined immediate memory performance on a wide range of simple and complex tasks and related it to verbal ability. They have found evidence for individual differences in general working memory capacity as being distinct from differences in STM mediated by rehearsal. Engle, Cantor and Carullo (1992) argue that working memory reflect a general capacity to hold information in LTM activated during processing. Just and Carpenter (1992) hold a similar view with the exception that the capacity limitation is specific to language information. According to our proposal superior text comprehension reflect superior skill in encoding information in LTM thus allowing a larger amount of information to remain accessible by means of sues in ST-WM. The LT-WM and capacity accounts differ in their predictions about storage in LTM and individual differences in factors facilitating processing and storage in LTM.

According to capacity theories the maintenance of words in the reading-span task is solely based on activation (ST-WM). However, subsequent research has accumulated substantial evidence implicating storage in LTM during the reading-span task. Masson and J. A. Miller (1983) found that, in cued recall, other words in the sentences were equally good as predictors of reading comprehension as the final words. Baddeley (1986) found similar correlations between comprehension ability and a modified reading-span test in which subjects were told only after the end of presentation what type of information they had to recall. In this version of the reading-span task, the subjects could not anticipate what information would be requested and therefore had to maintain a great deal of information to perform well. The storage of large amounts of information is consistent only with storage in LTM.

Furthermore, the ability to remember sentences is related to verbal ability and the ability to comprehend text. Masson and J. A. Miller (1983) found that delayed testing of recognition memory for explicit and inferred statements from a paragraph were both as highly related to ability to comprehend text as to the reading span scores. Ericsson and Karat (Ericsson & Chase, 1982) found that memory span for words in sentences was highly correlated with a test of verbal ability and that during an incidental memory test at the end of the test session subjects could recall close to 80 percent of the sentences when cued with a unique word from each sentence. Carpenter and Just (1989) have also found a higher memory span for sentences as a function of their subjects' reading span. More recently Cantor and Engle (1993) let two groups of subjects with high and low capacity of WM, respectively, memorize unrelated (Exp. 1) and thematically-related (Exp. 2) sentences. For unrelated sentences they found the typical fan effect for recognition judgments, but the slope of the fan was higher for the low-WM group than the high-WM group. Most interestingly the relation between verbal ability and WM capacity disappeared when the effects of the slope of the fan was controlled. For the thematically-related sentences the low-WM group still revealed a typical fan effect, whereas the high-WM group showed a negative fan, that is the more statements related to a concept the faster the judgments. These findings show simply that the presented sentences are encoded in LTM differently for the two groups. The low-WM group appears to encode the sentences in isolation or in small groups of thematically related ones, whereas the high-WM group is able to attain a more integrated representation of the sentences, especially for the thematically-related sentences. Difficulties to suppress irrelevant information during text comprehension, which have also been observed for older subjects in research on aging (Hasher & Zacks, 1988), could be easily accounted for in terms differences in ability to form integrated representations and situational models of the text. Other differences in encoding processes between high-and low-WM subjects were found by Engle et al.

(1992). They found that the high-WM subjects allocated their study time better under self-paced study conditions than low-WM subjects.

Text comprehension is closely related to verbal ability, which is often measured by tests of vocabulary and language use (grammar). Test of vocabulary and word meanings are correlated with text comprehension as well as reading span (Dixon, LeFevre, & Twilley, 1988). Reading span uniquely predicts text comprehension, even when the influence of word knowledge is statistically controlled. However, knowledge about words is but one of many aspects of skilled readers' knowledge about language and discourse. Reading span is also correlated with the ability to make inferences, although if one partials out subjects' performance on questions asking for information stated explicitly in the text, this relation is no longer significant (Dixon et al., 1988; Masson & J. A. Miller, 1983). Singer, Andrusiak, Reisdorf, and Black (1991) have shown, however, that this correlation depends on the type of inference. Subjects are likely to make bridging inferences as an integral part of constructing a textbase because these inferences are necessary to make the textbase coherent. The correlation of bridging inferences with reading span remains significant even after explicit memory is partialled out, at least when the memory load is substantial (premises are separated by three intervening sentences). On the other hand, subjects are likely to make deductive inferences in response to subsequent test questions, rather than during the reading of the text. Deductive inferences are independent of working memory capacity once explicit memory is accounted for.

Some of the best evidence for our notion of LT-WM and against inherent individual differences in the capacity of temporary working memory comes from research that systematically varies both verbal ability and relevant domain knowledge and studies their effect on text comprehension. Recht and Leslie (1988) selected four groups of subjects on the basis of their reading ability (high and low) and their knowledge about baseball (high and low). Wolfgang Schneider, Körkel, and Weinert (1989) similarly selected four groups based on aptitude (high and low IQ) and knowledge about soccer (high and low). Using a similar factorial design Walker (1987) varied general aptitude with knowledge about baseball. All three studies found that memory and comprehension of texts describing events in soccer or baseball were influenced only by the amount of knowledge (high-low). There was no evidence for a main effect of or an interaction with IQ or reading ability. Hence, students with low reading ability and expert knowledge clearly outperformed students with high reading ability and little knowledge.

In a more recent study Yekovich, Walker, Ogle and Thompson (1990) compared two groups of students of low verbal ability who differed in their knowledge about football (high/low). These students read both texts used in standard tests of text comprehension and texts about football, which were constructed to have structures similar to those in the corresponding standard texts. Comprehension of the texts was found to depend on an interaction between type of text and knowledge level. Students with high knowledge showed better comprehension of the football texts than they did the standard texts, and there was no reliable difference for the students with low knowledge. The largest differences in comprehension corresponded to the highest level of integration and generation of thematic inferences. Yekovich et al. (1990) argued that enriched knowledge about football allowed the students with high-knowledge to generate these inferences in the normal course of comprehending the football texts, as proposed by Kintsch (1988). They offered further support for that claim in a subsequent study by showing that students with a high level of knowledge about football (regardless of high or low verbal ability) could readily generate fluent thematic inferences when they are asked to comment concurrently on a football game.

The high correlations between text comprehension and, on the one hand, measures of long-term memory for texts and sentences and tests of language knowledge (vocabulary and grammar) are consistent with the assertion that text comprehension is an acquired skill. An important aspect of this acquired skill is storage of an integrated representation of the previous text in LTM. The storage itself must be rapid and accurate, and it must allow efficient retrieval of this information whenever needed. We suggest that Daneman and Carpenter's (1980) reading span measures this ability to store and later retrieve information about preceding sentences from LTM. Thus, what we are dealing with in the studies we have reviewed is not maintenance of temporary information in working memory, but skilled readers' ability to access LTM from retrieval cues held in the active portion of working memory.

In sum, Engle, Cantor and Carullo (1992) entertain the possibility that their results on working memory capacity is mediated by Chase and Ericsson's (1982) retrieval structures, but suggest that "there is a limitation either in the number of contexts or the number of links resulting from the limitation in total activation" (p. 990). Just and Carpenter (1992), however, have adduced a great deal of other experimental evidence in favor of their capacity theory of comprehension, in addition to the reading-span results. This theory can take two forms. In the form these investigators favor, the total capacity of working memory varies among individuals. A large capacity makes a good reader because it enables that reader to store more information in working memory during reading. In another version, which fits their data equally well, the total capacity does not differ between good and poor readers, but the processing efficiency of good readers is assumed to be higher, so that their effective working memory capacity is enlarged because they can use their resources better. Our proposal for LT-WM can do without the somewhat slippery notion of cognitive resources altogether. What is limited is merely the transient portion of working memory. Good readers perform better because their superior comprehension strategies result in the construction of better retrieval schemata. All the data Just and Carpenter (1992) report can readily be reinterpreted in this way. For instance, there is no need to claim that only high-span readers have the capacity to take pragmatics into account. Instead, it may be the case that skilled (hence high-span) readers have available sophisticated sentence parsing strategies based on pragmatic information that poor, low-span readers lack. This claim is supported by some very recent research by MacDonald and Pearlmutter (1993), who re-examined a finding by MacDonald, Just and Carpenter (1992) showing that high-span subjects took longer to read temporarily ambiguous sentences than low-span subjects. In contrast to Just and Carpenter's (1992) account in terms of capacity theory MacDonald and Pearlmutter (1993) show that only high-span subjects have sufficient knowledge of language to be sensitive to the probabilistic constraints guiding the ambiguity resolution. These findings and others that we reviewed earlier seriously question Just and Carpenter's (1992) assumption that good and poor readers perform the same operations, but that some have more room in working memory than others. In comparison, our proposal for LT-WM emphasizes good readers' use of more sophisticated, more complex comprehension strategies - procedures for the construction of mental representations - that result in the generation of more extensive retrieval structures and hence a larger effective working memory. This interpretation has two advantages. It is parsimonious (no recourse is necessary to ill-specified mental resources that take up cognitive capacity). More important, the kind of retrieval structures that are being built in text comprehension can be specified in some detail, as we show in the next section.

The Construction of an LTM Representation from Text: The Construction-Integration Model

We now describe a model of comprehension that explains how the processes involved in comprehension result in the construction of retrieval structures and thereby create LT-WM. This model is Kintsch's (1988; 1992a; 1992b; in press-b; Kintsch & Welsch, 1991) construction-integration (CI) model, which extends the theory of discourse comprehension developed by Kintsch and van Dijk (1978) and van Dijk and Kintsch (1983) through hypotheses about the activation and use of knowledge in comprehension. We can only sketch the principal features of the model here while focusing on the memory aspects of the theory.

The CI-model

The CI-model is a computational model of discourse comprehension. From the text it receives, it constructs mental representations which serve as the basis for free recall, question answering, priming effects, and other text-related behaviors. The operations used in this construction process simulate important aspects of human comprehension. The representations generated can be regarded as hypotheses about the nature of memory representations of texts in human readers.

The model uses as its input not the text itself but a hard-coded semantic (propositional) representation of the text. That is, it does not deal with problems of sentence parsing. It is concerned with how the individual meaning elements (propositions) are put together to form a coherent representation of the text as a whole. The basic assumption is that the rules for doing so are general, weak, relatively insensitive to the context, and inexact, so that the representations that are generated are full of redundancies, irrelevant information, and contradictions. Hence, a process for satisfying contextual constraints is needed - the integration process - to strengthen those elements of the representation that fit together and to weaken or reject those that do not.

Text comprehension is a sequential process. As each meaning element is generated (by means of the weak, approximate rules previously mentioned), it is integrated with the previous elements that are still held in the focus of attention. At sentence boundaries (or, if the sentence is too long, a suitable phrase boundary) the propositional network that has been generated is dropped from the focus of attention, though it remains available in LTM. Thus, what is stored in LTM is fully interpreted, contextually integrated text representations. As the reader proceeds to the next sentence, some of the text elements being processed are be linked with earlier portions of the text stored in LTM and thus serve as retrieval cues for these portions, creating a LT-WM. The mental representation of the text generated by the reader in accordance with the structure of the text thus comes to serve as a retrieval structure.

Access to Episodic Text Memory during Comprehension

Successful comprehension implies that the mental representation of the text is coherent, both at the level of the macrostructure of the text and at the local level. Texts are not always written in such a way, however, that a coherent structure can be generated, because the amount of text that can be held in the focus of attention at any point in the process is strictly limited (typically, that amount is the current sentence). It is therefore frequently the case that items from the episodic text memory under construction must be reinstated in the focus of attention to ensure the coherence of the memory representation (since operations of any kind can be performed only on material held in working memory). If the episodic text structure that has been generated is coherent, text elements currently in the focus of attention provide access to these structures. The elements that provide these links are the ones that provide for the linguistic coherence of the text: anaphoric and cataphoric elements in the episodic model of the current text, generic lexical knowledge, as well as contextual features. In a discussion of mental coherence, Givón (in press) emphasizes six classes of coherence elements: referents, temporality, aspectuality, modality/mood, location, and action/script. Hence any proposition located in the previous text and linked to these cues is directly available for reinstatement in the focus of attention through a single, 400-ms retrieval operation. The text structures that are generated in comprehension thus become the retrieval structures of LT-WM.

These structures can be quite complex, consisting partly of the propositional network derived form the text (the textbase), partly of associated knowledge. Situation models are not necessarily of a propositional nature. Imagery may be involved that integrates the text and the reader's domain knowledge and supports and supplements the information given by the text with relevant general knowledge or personal experience. The episodic text memory is therefore a very rich structure that connects text and knowledge as well as personal experience in many ways. As a consequence, a very large amount of information becomes potentially available in LT-WM through this retrieval structure that has been generated incidentally, as an integral part of text comprehension.

To help a reader construct a coherent mental representation of a text or discourse, speakers or writers insert in their texts syntactic cues that serve as processing instructions to the reader. Givón (in press) has described various types of anaphoric cues that indicate the rough location in the mental text structure of a prior referent. For instance, if the prior referent is still activated (usually that means that it was introduced not more than a clause back), the English language indicates a recent mention through zero anaphora or unstressed pronouns (e.g., *...He circled it warily as a wolf, [0] studying it from all angles,...*). At the other extreme, a definite noun phrase with a modifier is used as a long distance anaphora (e.g., *and when finally he stopped within a dozen feet of the dead man,...*)(See Footnote 2). In fact, language possesses a variety of graded syntactic devices to indicate to the processor just where in the text information that is to be reactivated is located. These are not only anaphoric cues but also cataphoric cues (e.g., referents marked with an indefinite "this" will recur as a central concept in the subsequent text). Syntax also instructs the reader when not to look for prior links but to start a new thematic unit by means of a variety of switching devices. For example, a plain "and" signals continuity (only 16% of the occurrences of "and" were associated with switches in topic in Givón's corpus), whereas an "And" following a period signals the beginning of a new thematic unit (100% switches). Thus, the cues present in a reader's focus of attention not only make possible retrieval from LT-WM but also indicate to the reader when to attempt such retrieval and when not to.

If the text comprehension processes fail to generate a coherent text representation, for example when the text is difficult or poorly written or when the reader lacks the domain knowledge for an adequate situational understanding, the retrieval structures that provide access to this large amount of information are not in place, or rather, are incomplete, and reinstatements may involve time- and resource-consuming searches (e.g., J. R. Miller & Kintsch, 1980). In such cases, the reader must first generate an appropriate retrieval cue, which can be a difficult problem-solving task in itself.

The textbases and situation models that the CI-simulation generates are the retrieval structures necessary for creating LT-WM. Other theories of discourse comprehension are not very different from the CI-model in this respect, however. There may be differences in the details concerning how these structures are generated and what their primary characteristics are, but in principle various kinds of structures would be capable of creating an LT-WM in discourse comprehension. For instance, both Carpenter and Just's READER model (Just & Carpenter, 1987, Chapter 9) and Gernsbacher's structure building approach (Gernsbacher, 1990) yield structures that could support retrieval from LT-WM in ways not very different from those described here. Even the structures generated by computational linguistics theories (e.g., Grosz & Sidner, 1986) could, in principle, serve this function. At some general level, comprehension is simply structure building, and in order for this large structure to be built sequentially in LTM relevant portions of it must remain accessible. This is what LT-WM enables readers to do in comprehending text.

Associative Activation of Knowledge during Comprehension

The episodic text structure is formed during comprehension and provides the retrieval structures needed to access prior portions of a text. The reader's already existing domain knowledge provides another source of retrieval structures with which to access relevant knowledge necessary for full understanding of the text.

In the CI-model knowledge is represented as an associative network. Lexical items as well as propositions encountered in a text associatively activate a few of their neighbors in this network. This associative activation does not take context into account, except that compound cues also can retrieve associated knowledge (Ratcliff & McKoon, 1988).

Consider as an example the following sentence pair (previously analyzed by J. R. Miller & Kintsch, 1980): "Eva Benassi was dying. She recovered from illness when her nurse prayed to the bishop". The first sentence may retrieve "funeral," "she recovered from illness" may retrieve "health," and "her nurse prayed to the bishop" may retrieve "religion" by themselves; but all three items together probably would retrieve "miracle." Neither "funeral," "health," or "religion" would be particularly context-appropriate in this case and would be suppressed by the contextual integration process; but "miracle" is just the right piece of knowledge. In a network consisting only of these nodes and starting out with equal activation for the three text proposition and no activation for the retrieved knowledge items, "miracle" becomes the most highly activated proposition - stronger than the text propositions themselves, whereas the other three knowledge items end up with only weak activation values. The model has, thus, inferred the topic of the passage.

Just as it is often necessary to reinstate textual information in the focus of attention during comprehension, it is also often necessary to retrieve general knowledge (or personal experiences) that is needed for the interpretation of the text (the construction of the situation model). The role that knowledge inferences play in text comprehension is varied and important (see Kintsch, 1993). Our concern here is solely with the memory retrieval requirements in inferencing. The resource demands of inferences in text comprehension differ widely. Some are quite automatic, whereas others may require a great deal of directed problem-solving activity.

Consider the contrast between the following two sentence pairs: "John's car broke down. The motor just quit." versus "An abnormally low amount of hydrocele was found. The spermatic cord appeared quite dry." A bridging inference is required in both cases, but the inference is trivial in the first case: The knowledge that cars have motors and that the motor quitting is sufficient cause for a car to break down is readily available; "car" and "motor" are effective retrieval cues that bring the relevant parts of LTM into working memory, so that a coherence problem does not even arise. But "hydrocele" and "spermatic cord" retrieve nothing for most readers. Either the sentence pair remains incoherent, or the reader engages in deliberate, conscious inference processes, hypothesizing, for instance, that because the spermatic cord was dry, it might be the place where a substance called hydrocele was low.

Domain knowledge plays a large role in text comprehension and memory (e.g., Afflerbach, 1990; Bransford & Johnson, 1972; Dooling & Lachman, 1971; Moravcsik & Kintsch, 1993; Recht & Leslie, 1988; Wolfgang Schneider et al., 1989; Spilich, Vesonder, Chiesi, & Voss, 1979). Domain knowledge provides the retrieval structures that give readers direct access to the information they need when they need it. Given a richly interconnected knowledge net, the retrieval cues in the focus of attention can access and retrieve a large amount of information. For instance, Kintsch and Keenan (1973) gave subjects sentence pairs such as the following to read: "A burning cigarette was carelessly discarded. The fire destroyed many acres of virgin forest," or sentence pairs in which the first sentence contained the explicit information that the "cigarette started a fire." Subjects then verified test sentences such as "A discarded cigarette started a fire." Reaction times were 400 ms (in another experiment with longer texts, 500 ms) faster when subjects had read the explicit statements than when they had to make the bridging inference on their own. On the other hand, when the test question was delayed for 20 m, there was no difference between conditions, presumably because retrieval from LTM was involved in both cases. Apparently, the retrieval of the episodic text structure also retrieved the associated knowledge about the general world, so that the information that burning cigarettes may cause forest fires was available in working memory, whether stated explicitly in the text or not. The 400-500 ms retrieval times are in good agreement with other estimates of retrieval times for LT-WM.

The Role of the Short-term Memory Buffer in Comprehension

In the Kintsch and van Dijk (1978) model and its successors, a limited-capacity, STM buffer plays an important role in comprehension. By maintaining information in working memory, the STM buffer facilitates the construction of a coherent textbase. A small number of propositions are selected at the end of each processing cycle and carried over in a buffer to be reprocessed with the input propositions from the next processing cycle.

The size of the buffer in various applications of the Kintsch and van Dijk (1978) model has been estimated as between one and four propositions (Kintsch & van Dijk, 1978; J. R. Miller & Kintsch, 1980; Spilich, 1983; Spilich et al., 1979). This estimate agrees well with other estimates of STM capacity. If all resources can be used for storage, as in a memory-span test, about seven to nine chunks can be retained (G. A. Miller, 1956). If, on the other hand, subjects learn a word list for later free recall, a task for which most resources are devoted to encoding of information in LTM rather than short-term storage, only about two items are reproduced from STM (Glanzer & Razel, 1974). The model requires the specification of a strategy for selecting propositions to be kept in the short-term buffer. Several successful strategies have been proposed and evaluated empirically (for reviews see Fletcher and Bloom, 1988; Kintsch, 1992a.)

The data cited in support of the short-term buffer are readily accounted for by LT-WM. Instead of assuming that a number of propositions are stored in a short-term buffer we propose that a comparable function is performed by the elements in STM that provide context and access to the relevant portions of the structure of text in LTM. We have already shown how this model can account for selective and rapid access to important propositions and their arguments, because they are part of the LTM structure that encodes the current situation model kept accessible by elements in STM. Given that LT-WM allows readers to maintain access to the important propositions at higher levels in the text hierarchy, it can account for the widely documented "levels effect" in text recall: The more superordinate propositions in a text are recalled better than subordinate propositions are (Kintsch, 1974; Meyer, 1975.)

An account of these phenomena in terms of LT-WM has, in addition, the advantage that it is not necessary to propose a separate strategy for selection of propositions for the short-term buffer. The comprehension process encodes new information into the structure of the text in LTM, and accessibility of propositions and other information is an indirect consequence of comprehension.

Summary

During fluent reading of well-written texts, mental representations of successive sentences are generated in ST-WM. Elements of that representation are linked both to parts of the previously constructed text representation (the episodic text memory), which is already stored in LTM, and to the reader's knowledge. This linkage creates a LT-WM that provides direct access to relevant parts of these structures from the cues available in STM. Once the reading of the text is completed, tests of comprehension and recall reflect the representation of the text in LTM.

Our model of working memory in text comprehension is distinguished from alternative models based on transient activation of information in STM by the central role of storage of accessible information in LTM. Consistent with our model, subjects' reading can be completely disrupted for over 30 s with no observable impairment of subsequent text comprehension. The observed increases in reading time, which are restricted to the first sentence after reading is resumed, occur because elements of the text structure have to be retrieved from LTM in order to reinstate the information available in STM before the interruption.

The construction of an integrated representation of a text in LTM is a skilled activity that requires prerequisite knowledge as well as encoding skills if an individual is to be able to successfully anticipate future retrieval demands. Differences in knowledge about the general topic of the text have been repeatedly shown to influence memory as well as comprehension of a text. Lack of prerequisite knowledge impairs both encoding and storage in LTM and the ability to generate the inferences needed to create an integrated representation The empirical evidence on individual differences in comprehension of standard texts is consistent with differential ability to encode information in LTM in a form that allows subsequent reliable access when this information is referenced or relevant. Differences in domain knowledge and knowledge of different genres of text are correlated with successful encoding abilities. Finally, the CI-model explains how LT-WM is involved in the construction of mental representations during text comprehension.

Memory in Problem Solving and Other Cognitive Activities

It is relatively easy to extend our proposal for LT-WM to most laboratory studies of problem solving and other types of complex cognitive activities. Many findings (discussed in the introduction) can be adequately accounted for by the classical model, in which cognitive processes are represented by a sequence of states of information in STM. In our view, however, the sufficiency of the classical model of working memory in accounting for many results in studies of problem solving and other complex activities is due to careful selection of laboratory tasks and stimuli to elicit these cognitive processes. Experimental psychologists have traditionally attempted to design laboratory tasks and situations that enable them to study a particular mental function in its purest form. Hence they have tried to minimize the role of memory in their studies of problem solving, thinking, reasoning, and decision making. In earlier studies, researchers tended to select task domains that were unfamiliar to subjects and hoped thereby to eliminate the influence of specific, pre-existing knowledge on the cognitive processes under observation. All the information necessary to perform the tasks therefore had to be presented to the subjects. Investigators alleviated the demand for memory storage in most studies by letting the information remain perceptually available during the task. Newell and Simon (1972) referred to perceptually available information as being accessible from external memory.

In addition, several characteristics of these unfamiliar tasks make use of LT-WM for storage difficult if not impossible. By definition, the stimuli used in an unfamiliar task do not correspond to pre-existing knowledge, which would have facilitated encoding and storage in LTM. Furthermore, the stimuli are often combinations of independent features that would maximize interference for encoding into and retrieval from LTM of stimuli subjects have seen previously. Finally, when subjects are unfamiliar with a task, they do not know what aspects of the information are important. Any efforts subjects would otherwise make to selectively store information with appropriate retrieval cues are therefore virtually impossible.

As subjects gain extensive experience with an initially unfamiliar task, they acquire a good deal of specific knowledge and develop procedures for generating answers and solutions. Hence, with extensive experience in a given task or domain, subjects can overcome the obstacles for using storage in LTM. Consistent with the superior memory of experts, reviewed in the introduction, subjects can extend working memory by acquiring memory skills for specific types of skilled cognitive activities.

Many types of skilled performance involve specific activities that can be easily distinguished from other everyday activities. In contrast, text comprehension involves language and comprehension skills that are closely related to most school and everyday activities. For this reason the amount of experience and improvement in comprehension are very hard to measure and distinguish from improvements due simply to maturation. Skilled performance therefore offers better opportunities to study the emergence of extended working memory as a function of practice and experience. It is possible to investigate working memory capacity and its mediating mechanisms in expert performance by studying individuals who differ in their level of performance in a domain.

We will begin our review of skilled performance by examining some simple skills in tasks for which investigators can dramatically increase demands on memory by eliminating external storage and requiring subjects to perform the tasks mentally. Then we consider increasingly complex skills and compare the performance of experts and novices in several domains.

Mental Calculation

Calculation in mathematics is taught in school systems all over the world, and the methods and strategies that guide cognitive processes in this domain are well understood. For mathematical problems students are typically allowed to use paper and pencil, a slide rule, an abacus, or in these times an electronic calculator to externalize storage of intermediate products and results. By restricting access to external storage and forcing subjects to carry out the calculation mentally, investigators can create an interesting case for which both the cognitive processes and demands on working memory can be explicated. The demands on working memory in mental calculation differ depending on the type of problem and mathematical operation involved. Here we will consider only addition and multiplication.

Mental Addition

For most adults, adding multi-digit numbers with paper and pencil is easy. When they have to keep all the information in memory, however, they often produce incorrect answers. Hitch (1978) found that subjects adding numbers such as 325 + 46 often made errors. These errors were not random but could be predicted by the length of time subjects had to keep the digits generated for the answer in STM before reporting the answer. Untrained subjects apparently kept intermediate products and results primarily in transient STM, which decays rapidly unless it is rehearsed.

Studies of skilled subjects performing complex addition have been conducted primarily in Japan with experienced abacus operators. The task is typically to add sequentially presented numbers up to nine digits long on a perceptually available abacus. The current subtotal is represented on the abacus, and each new presented number is added to it. Hence, there is no benefit from storing any subtotals except the most recent one. Of particular relevance to working memory, the abacus operators who are most frequently studied are able to perform the addition task mentally and report relying on a mental abacus. Calculation with a mental abacus is by no means an automatic consequence of extensive experience with a physical abacus, and most investigators (Stigler, 1984) view mental calculation as a separate skill acquired through practice specifically on mental problems. Acquiring a mental abacus is a slow process that starts with the ability to represent 3 or 4 digits. The rule of thumb is that expanding the mental abacus by one additional digit takes about one year of deliberate effort (Hatano & Osawa, 1983).

Calculation with a mental abacus leads to the same type of errors as working with a real abacus does, and these errors are markedly different from the errors in mental calculation produced by American subjects (Stigler, 1984). Furthermore, Stigler (1984) found that skilled abacus calculators appear to have access to distinct descriptions of intermediate states of the abacus even during mental addition. When the calculators were asked to monitor for a specific configuration of beads of the abacus during a calculation, their pattern of reaction times is similar for both mental calculation and calculation with an actual abacus.

Several investigators have studied differences in memory capacity and representation as a function of level of skill in mental abacus calculation. Subjects with higher skill levels have a larger memory capacity for digits, as measured by the digit span (Hatano, Amaiwa, & Shimizu, 1987). Elite-level subjects have exceptional digit spans between 12 and 16 digits, but their memory capacity for other types of materials is within the normal range (Hatano & Osawa, 1983). Subjects with higher skill levels show evidence of a different type of memory encoding of presented digits. Articulatory suppression has considerable effects on the memory performance of less skilled subjects, but no effect is observed for skilled subjects (Hatano et al., 1987; Hatano & Osawa, 1983). The effect of a concurrent visual-spatial task is greater for elite than for less skilled subjects (Hatano & Osawa, 1983). Experts' memory for digits is affected more by exposure to irrelevant abacus configurations than exposure to lists of digits during a retention interval, whereas less skilled subjects show the reverse pattern of interference (Hatta, Hirose, Ikeda, & Fukuhara, 1989).

The accessibility of presented digits in memory also changes as a function of skill level from a sequential organization to a more flexible representation that allows order-independent access. When experts are instructed to report presented digits in reverse order (backward digit span), their digit span is not affected. The performance of less skilled subjects is reliably reduced in this task (Hatano et al., 1987). Cued recall of digits represented within the mental abacus elicited no increases in experts' reaction time as a function of serial position, whereas less skilled subjects showed the typical linear increase in reaction time associated with a sequential representation of stored digits (Hishitani, 1990).

There is controversy over how the mental abacus should be characterized in terms of working memory. Hatano and Osawa (1983) performed one experiment that they claimed ruled out storage in LTM. Two elite subjects were asked to maintain ten digits for 30 s for ten consecutive lists, after which they were given a surprise recall and recognition test. One of the subjects could accurately recall the last list, but otherwise these subjects exhibited no evidence for memory of the presented lists. This result is not inconsistent, however, with notion that the mental abacus is an instance of LT-WM.

The demands on working memory in mental abacus calculation require that during addition of a new number, the current subtotal be updated digit by digit in an independent manner. Hence, the most plausible implementation of LT-WM would be storage or activation of an encoding of the locations of one or more beads for each of the relevant positions on the abacus. With each operation (addition of individual digits) a new configuration of locations would be linked to that position. Because the same spatial position cue is continuously reused, interference with a previously stored (activated) memory process should be a major problem. Indeed, it explains the poor memory for previously seen lists of digits that Hatano and Osawa (1983) observed in their elite subjects. Our account in terms of LT-WM is superior to alternatives involving ST-WM in that it easily accommodates the restriction of the superior memory capacity to only numbers and the long period of relevant training necessary to attain this domain-specific form of extended working memory.

Mental Multiplication

The problem of multiplying two multi-digit numbers is a natural extension of addition because addition is a significant component of multiplication. Similar to addition problems, multiplication problems offer most educated individuals no real difficulty when paper and pencil are available for external storage. When multiplication has to be completed without any external memory aid, it becomes a most challenging task. Unlike mental addition of a list of sequentially presented numbers, where only a single subtotal has to be kept in memory, mental multiplication requires intermittent access to the numbers of the original problem and to calculated intermediate products during most of the calculation process, which can take up to several minutes to complete.

Like mental abacus calculation, mental multiplication is a distinct skill that requires specific practice beyond extended experience of multiplication with paper and pencil. All of the most outstanding mental calculators (Hunter, 1962; Jensen, 1990; S. B. Smith, 1983) spent many years practicing before achieving their superior performance. Large individual differences in mental calculation among college students were found to be closely related to prior amounts of specific experience (Dansereau, 1969). In an extended training study, Staszewski (1988b) showed that after hundreds of hours of practice, ordinary college students could improve their mental calculation performance by a factor of 5 and match the performance of a professional performer on arbitrary multiplication problems in a particular format used in practice.

Research on outstanding mental calculators has shown that most of them have accumulated a vast body of knowledge about mathematics and methods of calculation (Hunter, 1962; S. B. Smith, 1983), which would be very difficult and time consuming to describe. Hence investigators have concentrated on examining the basic capacities and abilities of these mental calculators as measured by standard psychometric tests. The pattern of results is remarkably clear: mental calculators have exceptional memory for digits, but their memory for other types of material is within the normal range (Ericsson, 1985; Jensen, 1990). Although their digit span is typically outside of the normal range, however, it is still below 18. Thus, a mental calculator who has decades of intense experience with numerical calculation only attains a level of performance on the digit-span task that normal college students can attain after less than 50 hours of specific practice (Ericsson, 1985), which is only a minute fraction fo the 10,000 hours estimated to be necessary to attain high levels of expert performance (See Ericsson, Krampe and Tesch-Römer, 1993, for a review).

Most systematic studies of working memory during mental multiplication have been restricted to college students. In an early study of a single subject, Dansereau and Gregg (1966) found that this subject's time to complete mental multiplications could be fairly well predicted by the number of operations necessary for solving these problems with the paper and pencil method. In a remarkable study Dansereau (1969) went beyond the obviously incorrect assumption that all operations would take the same amount of time and tried to estimate the duration of each operation. Using estimates from think-aloud protocols and performance on several specifically designed tasks, he estimated storage and retrieval time for LTM and STM. To simulate the performance of his subjects, however, he also had to postulate an intermediate-term memory store with its own time parameters. This memory store had temporal characteristics for storage and retrieval comparable to our estimates for LT-WM. Most importantly, individual differences in mental-calculation skill were associated with large differences in the speed of storage in and retrieval from this store. Consistent with the acquired nature of LT-WM, the memory of the best and most experienced calculators was the most efficient.

In an extended training study Staszewski (1988b) was able to trace the changes in the processes his subjects used as they dramatically improved their performance of mental calculations. Subjects developed strategies to encode intermediate results for efficient retrieval. With practice they needed less time to store information in LTM and were better able to later recall most of the multiplication problems they had encountered during a session. These findings clearly demonstrate storage in LTM during mental calculation. Furthermore, Staszewski found evidence for the specificity of LT-WM. Throughout training his subjects had solved problems in which a 2-digit, 3-digit, 4-digit or 5-digit number is multiplied by a 2-digit number and only two intermediate products are generated. With these problems one of his subjects matched the solution times of an experienced mental calculator. When tested on problems for which a 3-digit number is multiplied by another 3-digit number, which requires calculation of three intermediate products, this trained subject's solution times were twice those of the experienced mental calculator. A likely source of the difficulty of transfer to the new problem type concerns the inadequacy of the old retrieval structure to accommodate efficient and reliable storage of one additional intermediate result.

Further evidence for encoding in LT-WM during mental calculation was obtained by Chase (Chase & Ericsson, 1982). He analyzed a mental calculator who specialized in squaring 3-, 4-, and 5-digit numbers. This expert had discovered an algorithm for squaring that dramatically reduced the memory load; but even so, he had to store a small number of intermediate results without intervening rehearsal for a significant portion of the calculation period. Chase found that his subject used mnemonic methods to encode these intermediate results but did not use these methods for other temporary results. This result is particularly interesting in that is shows that an expert selects different encoding methods to fit the specific retrieval demands of different intermediate results within a given calculational procedure

In summary, working memory in mental addition and multiplication is consistent with the two types of encodings we proposed for LT-WM. Rapidly changing intermediate results in mental abacus calculation and during addition in mental multiplication are stored by activation of a code that links the result and its retrieval cue, whenever proactive interference limits the duration of reliable access. Intermediate results and information about the problem in mental multiplication must be deliberately stored in LTM for retrieval to be accurate and reliable in the absence of intervening access over long periods of time.

Other Everyday Activities

Most everyday activities are so habitual and adapted to their social and physical environments that it is difficult to assess their structure and demands on working memory. Every since Ebbinghaus (1885/1964), investigators have recognized the relation between knowledge, interest and accuracy of memory for new information in a domain. In this paper we have reviewed the evidence for superior memory in experts and individuals who know a great deal about a domain. A large body of experimental studies of memory shows that memory for a specific type of material improves with practice (Ericsson, 1985). Think-aloud protocols have revealed that individual differences in memory performance reflect differences in relevant knowledge (Gilhooley et al., 1988) and in encoding strategies, which can be improved by instruction (Thorndyke & Stasz, 1980). Some of the largest differences in memory performance are found for memory tests that capture naturally reoccurring situations. Many sports fans obtain information about the results of matches between sports teams as a list, often heard on the radio or TV. The amount of knowledge of soccer was found to be a powerful determinant ($r > 0.8$) of subjects' recall of newly presented scores for recent soccer matches (Morris, Tweedy, & Gruneberg, 1985). Even for lists of simulated soccer matches to which a score was randomly assigned, accuracy of recall increased with knowledge of soccer, but the strength of the relation was weaker than for actual soccer scores.

Further evidence of working memory in everyday activities can be found in waiters and waitresses, who repeatedly take drink or dinner orders from large parties of customers. Typically, these individuals write down the information especially for complex dinner orders, in a systematic fashion. Many experienced waiters and waitresses memorize the orders, however, because by doing so they can be more attentive to the customers and thereby earn a larger tip. While memorizing orders, they have to be willing to explain and describe menu items, accommodate customers who change their minds, and so forth. Hence, in real life, memory skills must be compatible with interruptions and many other concurrent activities that demand attention.

Superior memory for drink orders (Bennett, 1983) and dinner orders (Ericsson & Polson, 1988a; 1988b) in experienced waiters and waitresses has been studied with laboratory analogs of the real situation in restaurants. From interviews at the end of the experimental sessions, Bennett (1983) found that expert waitresses encoded drink orders by linking the order to the corresponding person and location at the table. At the time of recall they could retrieve these orders by using the table with its seating locations as direct cues for the drink orders or for the customers, where the cues linked to the customers in turn might provide access to their drink orders. The same kind of encoding methods were revealed in concurrent verbal reports (Ericsson & Simon, 1993) collected from skilled waiters and waitresses by Crutcher and Ericsson (Ericsson & Polson, 1988b). The exception to this pattern was an exceptional subject (JC), who could memorize dinner orders from up to 20 customers, an unusually high number. Ericsson and Polson (1988a, 1988b) interviewed JC and designed a series of experiments to study his memory performance. Consistent with subjects in other studies of memory skills, JC reported that he only gradually acquired his exceptional memory performance. He started with parties of small numbers of new customers and then moved on to increasingly larger parties over a period of several years. From think-aloud protocols recorded while he memorized dinner orders, Ericsson and Polson (1988a) found that JC did not memorize a dinner order by associating it directly to the customer. Each dinner order in the laboratory analog situation consisted of an entrée for a beef dish, a cooking temperature (e.g., medium rare), a starch (e.g., baked potato), and a salad dressing. JC had devised a coding system by means of which he encoded all items of a category, such as the starches, together in a pattern linked to the location at the table. JC could then exploit the frequent repetitions of the same starch for different customers, which caused interference and confusion for the other subjects in the memory experiments, could now be used by JC to encode regular patterns, such as rice, fries, fries, rice. JC encoded salad dressings by their first letters (e.g., B=blue cheese, O=oil and vinegar), which he could often encode again as words (e.g., BOOT), or abbreviations and acronyms (e.g. HBO). He “visualized” cooking temperatures on a scale from rare to well-done and formed spatial patterns, for example, "well done" down to "medium rare" back to "well done.” For parties of more than four or five customers JC could split the list of categories into sublists of four customers organized by location around the table in clock-wise order.

While thinking aloud during the study of dinner orders JC would access and verbalize previously presented items in a category in order to encode a pattern for the entire sublist of four items. An analysis of his study times for individual dinner orders showed a clear increase until the sublist of all four items had been encoded (Ericsson & Polson, 1988b). After this point a clear decrease in study time was observed for the fifth dinner order, followed by a gradual increase until the second sublist of four items had been completely presented. Furthermore, JC would recall the studied dinner orders by category with the items reported in a consistent clock-wise direction.

Ericsson and Polson (1988b) designed a series experiments to validate the organization of the retrieval structure inferred from JC's reports and performance. They systematically varied the order in which the items were presented and found the same order of recall regardless of the order in which the items had been presented. JC’s encoding of the presented information into the retrieval structure was further supported by an analysis of think-aloud protocols (Ericsson & Polson, 1988a). In two other experiments they also varied the type of material presented to JC for memorization. As long as the semantic structure of the new materials allowed JC to apply his old encoding methods and retrieval structure, he adjusted quickly to these changes. When conditions were changed so much that this retrieval structure could no longer be applied, however, JC's performance deteriorated and he resorted to the less efficient strategies used by other waiters.

JC's memory performance was based on storage in LTM (Ericsson & Polson, 1988a). When JC was unexpectedly tested on his memory for dinner orders taken from customers in the restaurant, where he worked, his recalled most of the information with high accuracy. His post-session recall of dinner orders presented during the testing in the laboratory was very good for the most recent list for all different lengths of lists. However, he also showed strong effects of retroactive interference. When dinner orders for two tables with the same number of customers were presented in a session he could only recall the most recent list of dinner orders with high accuracy.

Differences in memory performance and the capacity of working memory in many everyday activities implicate differences in encoding processes and storage in LTM. The best performance can be linked to skills acquired in response to naturally occurring situations that necessitate rapid storage of information. This performance is mediated by retrieval structures and efficient encoding methods.

Medical Expertise

Experts in medicine gather and integrate different types of information in order to determine the correct action in complex situations. Medical diagnosis is particularly relevant to issues of working memory because symptoms and information about a given patient are received piece by piece over an extended time. These conditions raise the issue of interim storage prior to generation of a correct diagnosis. Moreover, medical diagnosis has been extensively studied under controlled laboratory conditions because it has been possible to generate laboratory analogs of these real-life tasks with vignettes about actual patients in such a way that the superior diagnostic accuracy of medical experts is preserved.

Expertise in diagnosing medial pathologies in a specific area of medicine is acquired over a long period of time. Before students can enter medical school, they must acquire certain prerequisite knowledge about science. Studies in medical school last for around five years and are then followed by specialized training in a specific field of medicine. Medical experts have typically had an additional five to ten year of experience in treating patients in their area of specialization. Diagnostic accuracy increases monotonically as a function of training and experience (Patel & Groen, 1991; Schmidt and Boshuizen, 1993).

In a laboratory analog of the medical diagnosis task , subjects are either presented with a text describing a particular patient or given the same information sentence by sentence (Boshuizen & Schmidt, 1992; Patel & Arocha, 1993; Patel & Groen, 1986). The most salient similarity between normal text comprehension and medical diagnosis is that the end results should reflect an integrated and accurate representation of the presented information. However, the intermediate steps toward this goal differ. A well-written text guides the reader in building the correct representation by including appropriate rhetorical signals that emphasize important information. In contrast, the description of a patient simply lists information. Only at the point when the correct medical diagnosis is generated can final decisions about the importance of individual facts and the relations between them be made. If information is prematurely disregarded or incorrectly encoded in light of an early incorrect diagnostic hypothesis, it is difficult to recover and generate the correct diagnosis. It is thus necessary to limit encodings of encountered information to inferences that remain invariant across alternative diagnostic hypotheses.

Empirical evidence on diagnostic expertise is consistent with the acquisition of a retrieval structure that allow experts to encode basic medical facts about a patient into higher level diagnostic facts (Patel & Arocha, 1993) that can be used to access the correct diagnostic category and specific diagnosis. This hypothesis leads to a number of empirically testable predictions. First, the relevant information about the patient should remain accessible after its presentation. Second, experts should be relatively insensitive to the order in which information is presented and at the time of recall, reveal an order of recall that reflects the associated retrieval structure (as JC did with dinner orders).

A large number of studies have examined recall of information about patients as a function of medical expertise (See Patel & Groen, 1991 and Schmidt & Boshuizen, 1993, for reviews). When exposure to a medical case is limited, recall increases as a function of level of expertise. When more time is available to study the case, however, an inverse-U function typically relates recall and expertise. Recall by subjects with an intermediate level of expertise is higher than that of both less and more experienced subjects (Schmidt & Boshuizen, 1993). This surprising result appears to be due to two complementary factors. When students are asked to memorize information, their recall of all presented information increases compared to a condition in which they are asked to diagnose the patient and incidental recall is measured. Experts show the reverse pattern, and their incidental memory for the patient is higher when they are offering a diagnosis than when they are asked to memorize the information (G. R. Norman et al., 1989). Part of the problem is that the dependent variable measures recall of all the presented information, whereas adequate comprehension of the case reflects only the significant information.

A more refined analysis by Patel, Groen and their colleagues (see Groen & Patel, 1988, for a review) found support for this interpretation. Schmidt and Boshuzien (1993), however, found evidence for another related factor. From think-aloud studies of medical diagnosis, Boshuizen and Schmidt (1992) found that with increasing levels of expertise, subjects did not need to consult their biomedical knowledge to relate facts and symptoms of patients to diagnosis, but were able to use their clinical knowledge directly. In support of that hypothesis, Schmidt and Boshuzien (1993) were able to show that experts' free recall became more abstract and summary-like as their level of expertise increased. Fact recall was replaced by higher level statements that subsumed the specific facts. After extensive clinical experience, medical experts are able to acquire higher level concepts that can be induced from data on patients and allow for more effective reasoning about medical diagnosis. This representation allows medical experts to process information about typical patients in a bottom-up mode using forward-reasoning strategies (Patel & Groen, 1991) similar to normal text comprehension, as in Kintsch's (1988) construction integration model (Schmidt & Boshuizen, 1993).

It is difficult to determine how information about a patient is stored in memory because recall about a patient typically reflects the orderly organization in which the information was initially presented in the description of the patient. Analyses of subject's order of recall have suggested schemas for patient information organized by categories (Claessen & Boshuizen, 1985). Evidence for such an organization was obtained by Coughlin and Patel (1987), who presented experts and students with both a typically organized description of a patient and a version with the same sentences in scrambled order. Although subjects were given the same amount of time to study both types of descriptions, the diagnostic accuracy of experts was largely unaffected by scrambling and was higher than for the students.

A reanalysis of the order of recall for the scrambled texts (Groen & Patel, 1988) showed that all the experts and most of the students reorganized the presented information and recalled it in categories, as proposed by Claessen and Boshuizen (1985). A similar result was obtained by G. R. Norman et al. (1989) for laboratory-test results of patients. They compared recall by novices, students and experts for two different organizations of the laboratory test results -- one organized in meaningful, familiar categories and the other scrambled. Recall was unaffected when the order was scrambled, and the amount of recall increased monotonically with level of expertise. An analysis of the order of recall showed that the experts and to a lesser degree the medical students reproduced even the scrambled lists according to their appropriate conceptual categories.

Although the concept of LT-WM has not been explicitly considered in research on medical expertise, there is consistent evidence that experts store information about patients that they use to generate accurate diagnoses and that they can reliably recall upon request. Furthermore, intriguing evidence for a retrieval structure is provided by several studies in which subjects reorganized scrambled information into a semantically meaningful structure consistent with a patient schema. Finally, Patel and Groen and colleagues (Groen & Patel, 1988; Patel & Groen, 1991) have shown that experts can provide clear explanations for their diagnoses and fully account for all of the relevant presented information. It is reasonable to assume that a critical function of LT-WM in medical experts is not only to attain the correct diagnosis but to provide working memory support for reasoning about and evaluation of diagnostic alternatives.

Chess

Chess was the first domain of expertise to be systematically investigated (Chase & Simon, 1973; de Groot (1946/1978), and it has remained the most studied domain in cognitive psychology. It is thus not surprising that the most extensive evidence on LT-WM has been accumulated in this domain. Chess is an attractive domain to study because it has a very large number of active players and the better players have a chess rating, based on their performance in chess tournaments, that reliably measures their chess skill on an interval scale (Elo, 1986). As is the case for other types of expert performance it is difficult to study the chess experts' cognitive processes as they naturally unfold during chess games. De Groot (1946/1978) therefore devised a task that captures the essence of chess which is to select consistently the best move for a chess position. In his task, an unfamiliar mid-game chess position is presented and subjects select the best next move. Performance on the move-selection task is highly correlated with official chess ratings(Charness, 1991; de Groot, 1946/1978; Saariluoma, 1990).

By having chess experts think aloud while they select their move, de Groot (1946/1978) found evidence for extensive planning and evaluation of possible move sequences before the subjects announced the selected move. Transcripts of the protocols reveal that some of the best players (grand masters) were quickly able to identify the best potential moves, whereas other grand masters uncovered the best move during systematic exploration and evaluation (de Groot, 1946/1978). Hence, systematic evaluation appears necessary to guarantee highly accurate selection of moves. When de Groot analyzed the depth and other characteristics of very advanced chess players' search processes, he did not find any systematic differences related to differences in chess skill. In a subsequent study of chess players whose skill differed more widely, Charness (1981b) found a reliable relation between the maximum number of chess moves planned ahead (depth of search) and chess skill. Saariluoma (1991a) found that chess masters generated potential moves much faster and more fluently than novices in chess. Charness (1989) and Saariluoma (1990, 1992) have shown that the depth of planning during the selection of a move increases with chess skill up to the level of an advanced chess expert. The absence of further increases in depth of planning appears to be due not to lack of ability but to the emergence of a more sophisticated focus of evaluation and abstract planning (Saariluoma, 1990, 1992).

Planning during move selection imposes a major load on working memory. Saariluoma (1991a) has studied the effect on move generation from concurrent tasks that interfere with the articulatory loop and the visual-spatial scratch pad (Baddeley, 1986). He found no effect from articulatory suppression but did find a reliable effect from concurrent visual-spatial tasks (Saariluoma, 1991a). The effect did not differ for chess masters and novices, however, and the visual-spatial tasks may draw on central resources of perception and attention rather than selectively interfering with working memory.

The critical demand on working memory in skilled chess playing is the need to plan the consequences of long sequences of moves. Consistent with other skills that are acquired, the ability to plan increases during the first few years of serious chess study. Furthermore, the representation in working memory of planned chess positions reflects the characteristics of actual chess positions and allows chess players to uncover the strengths and weaknesses of these positions and to accurately evaluate and analyze them. Ericsson and Oliver (1984; Ericsson & Staszewski, 1989) proposed that the retrieval structures for chess position in memory corresponds to an actual chess board which allows access to each of the board's 64 squares. A chess position is represented as an integrated hierarchical structure relating the different pieces to each other, and all pieces are associated with their corresponding locations. Three kinds of evidence, discussed below, support the claims for such a retrieval structure. First, skilled chess players automatically encode the position of a piece. Second, for skilled players the location of a piece constitutes an effective retrieval cue. Third, experts are able to mentally manipulate and update their memory representation of the chess board.

In the laboratory a chess position is typically represented visually by chess pieces on a board or as a diagram, but it is possible to convey the same information as a list of all the pieces with their respective locations on the chess board. Saariluoma (1989, Exp. 2) found that after listening to a reading of a list of pieces with their locations in a regular chess position, chess masters were able to recall the position almost perfectly. Less skilled players were somewhat less accurate, recalling around 90% of the presented pieces. When the list of pieces was presented at a faster rate (2 s/piece), of the list of pieces accuracy of recall decreased, but chess masters were still 80-90% accurate. Saariluoma (1989, Exp. 1) found that recall was higher for ordered lists in which all chess pieces of the same color were presented together than for lists that presented pieces in a completely random order. With this type of presentation Saariluoma (1989) found that skilled chess players were able to encode and recall randomly arranged chess pieces and that the level of recall was closely related to chess skill.

Random chess positions provide particularly strong evidence that the ability to encode individual chess pieces into the retrieval structure is an acquired skill. At the same time several findings show that meaningful relations between chess pieces are also encoded and that integrated representations are formed for regular chess positions, but not for random positions. Recall of regular game positions is always much more accurate than recall of random positions. In one condition skilled chess players were presented with four positions, either four regular or four random, and then asked to recall them. These players could still recall the regular game positions well--chess maters' accuracy was around 60% --whereas recall of the random positions was below 10%. This finding suggests that meaningful chess positions can be integrated into distinct structures whereas random positions cannot.

A retrieval structure should also allow skilled players to rapidly retrieve select information in response to presented cues, in this case cues specifying a location on the chessboard. Ericsson and Oliver (1984; Ericsson & Staszewski, 1989) had a chess master memorize two different chess positions which required around 20 s of total study time. They then presented this subject with a cue specifying which one of the two positions and its particular location (square) on the chess board. The subject's task was to name the piece in that location or to say "Nothing" if that location was not occupied. This task was repeated for all locations of the two chessboards. In one of the conditions the locations of only one of the chess positions were repeatedly probed, and the speed of retrieval was fast. In another condition the cued chess position alternated unpredictably between the two chess positions, and the resulting retrieval time increased by around 2 s.

Ericsson and Oliver (1984) also compared retrieval of various types of information from a perceptually available chess position and from a memorized position. When the task was to count up the number of pieces in rows, columns, and other designated areas of the chessboard, the chess master was somewhat faster with a perceptually available position, but most interestingly, the speed of retrieval showed different patterns as a function of the number of squares that had to be considered. When the task was to retrieve the information from memory, retrieval time increased with the number of squares scanned, unlike retrieval time for perceptually available chess positions. Finally, Ericsson and Oliver found that the chess master could very rapidly extract the number of pieces of a given color that attacked a randomly selected square in a memorized position. In sum, these studies show that a chess master holds the information about a given chess position in highly accessible form and can readily produce information relevant to any of the squares upon demand.

The crucial test of a retrieval structure in chess is whether it allows skilled players to accurately represent dynamically changing board positions without external perceptual support. Chess masters must have such an ability, because they are able to play blindfold chess games. In blindfold chess players cannot see the board position and thus have to represent the current chess position in memory. Many master-level players are reported to be able to play blindfolded at close to their normal chess skill (Holding, 1985). To examine the ability of a chess master to mentally represent a chess game, Ericsson and Oliver (1984; Ericsson & Staszewski, 1989) presented the chess notation for one move on the CRT at a time for unfamiliar chess games without a perceptually available chessboard. After over 40 chess pieces had been moved, the chess master was tested by cued recall of the contents of all 64 squares of the chessboard. The chess master was able to recall information about chess position during the memory test with virtually perfect accuracy. His speed of retrieval from memory during the test was fast and only marginally slower than in another test condition when the current chess position was perceptually available during the presentation of the same set of questions.

In a series of studies with chess players at several different levels of skill, Saariluoma (1991b) presented the moves of actual chess games auditorily at the rate of one moved piece every 2 s and tested recall of the generated position after 30 and 50 pieces had been moved. Chess masters' recall was virtually perfect, chess experts' recall deteriorated with increased numbers of moves to around 40%, and novices were totally unable to perform the task. Saariluoma (1991b, Exp. 7) showed that under self-paced conditions of study a grand master could maintain 10 simultaneously presented blindfold chess games virtually without error. Chess masters and chess experts could also perform this task, but their accuracy of recall was reduced as a function of their level of skill.

In a couple of experiments Saariluoma (1991b) attempted to determine how chess positions are stored in working memory while subjects mentally construct the current chess position from a sequence of verbally presented moves. Concurrent articulatory suppression had no effect, a finding that ruled out the articulatory loop for storage. Concurrent imagery tasks degraded performance, but performance of the imagery tasks and other attention-demanding tasks during a pause in the presentation of moves had no effect. These results clearly implicate LT-WM in the storage of chess positions.

Research on planning and memory of chess positions offers some of the most compelling evidence for LT-WM. Saarilouma's (1991b) demonstration that a grand master could keep ten different games in during blind-fold chess is remarkable. However, George Koltanowski (1985) have several times played blind-fold chess against thirty or more opponents, where he won most of the games and the rest of games resulted in a draw. The achievements of grand masters in chess reflect memory and other types of skills acquired over ten years of deliberate practice (See Ericsson et al., 1993, for a review) and playing blind-fold chess against many opponents requires the acquisition of additional specialized skill (Koltanowski, 1985).

Problem Solving in Knowledge-Rich Domains

We conclude our review with activities experts perform to construct integrated representations in memory of a design, computer program, or text. In most of these cases individuals base their constructions on a very large body of relevant knowledge. Because of vast differences in the knowledge available to each individual, research on the design of computer programs, complex social problem solving, and composition of texts has predominantly focused on structural differences in novices' and experts' organization of cognitive processes. Across these three very different task domains, remarkably consistent differences between experts and novices have been found. Not surprisingly, novices tend to start working on a computer program or text without much, if any, planning and preparation. This is most striking for novice writers, who write their texts as a sequence of ideas generated from memory using the provided title and their most recently generated idea as cues. Bereiter and Scardamalia (1987) presented a large body of evidence showing that novice writers produce their texts from repeated probing of memory (knowledge telling) without any prior attempt to reorganize and transform this knowledge for the reader. In contrast, all types of experts tend to spend a substantial amount of time reviewing the problem, considering constraints, and decomposing the problem into a sequence of subproblems that can be solved independently or with minimal interaction.

Basically, the solution of a complex design problem involves mapping in detail the interactions between various parts of the design product. In a computer program, for example, all modules must use the same data representation, and the calculations of certain modules must be passed to other modules. Experts in software design generate complete models of the program at increasing levels of specificity before starting to produce the solution, that is, actually writing the code for the program (Jeffries, Turner, Polson, & Atwood, 1981). According to Jeffries et al. (1981), this design process is guided by design schemas available in experts' LTM. In a simpler task studied by Byrne (1977), skilled subjects generated a series of three-course meals for the same guests. In generating the components of the main and second courses, subjects systematically started with the type of protein or meat and then decided on the starch and vegetables, as a sequence that would facilitate the coordination of these components. Byrne (1977) even found that subjects would frequently generate the courses out of order and start with the main and second courses to make the selection of a coordinated first course easier.

In another kind of problem-solving experiment, Voss, Greene, Post, and Penner (1983) asked different types of experts and novices to describe how the Soviet Union could increase its agricultural production. Experts on the Soviet Union tended to begin by outlining the history of previous attempts to solve this problem and explored general solutions to increasing depths of specificity in a hierarchical manner. The hierarchical approach allowed experts to consider the general problem of capital investment and its observable implications for more specific problems, such as lack of repair parts and fertilizer. The novices, on the other hand, tended to enumerate these surface problems as separate solutions. Domain-specific knowledge about the Soviet Union was found to be critical, and expert political scientists without that knowledge were able to generate only the abstract issues without being able to proceed to detailed solutions. Experts in subjects outside of political science, such as chemistry, tended to resemble novices in both the content and structure of their solutions.

In contrast to the experts on the Soviet Union, expert writers are frequently not experts on the topics they are writing about. From thinking-aloud studies of writing, Hayes and Flower (1980) demonstrated that experienced writers tend to start by generating ideas and content and spend considerable time organizing this content before beginning to write. Hence, it appears that the experienced writers construct a well-formed organization for their ideas, whereas the domain-specific experts are able to draw on pre-existing cognitive structures.

The composition processes of serious poets and control subjects in an early think-aloud study by Patrick (1935) are similar in structure to those of expert and novice writers. Patrick found that the first phase of writing a poem was dominated by the generation of possible themes, ideas, and images, although she described this phase by the changes in ideas and themes and stressed the non-cumulative character of subjects' thoughts. As a theme emerged, the subjects started writing lines of the poems and toward the end of their work were occupied mostly with revisions or editing.

In summary, the results from analyses of experts' cognitive processes in tasks requiring extensive retrieval tell us something about the role of memory. Through the initial planning and design activity, the expert is able to hierarchically decompose the problem or product into subproblems or subactivities (Simon, 1973). In this way the expert simplifies the actual production of the solution or product, which would be taxing to working memory, and limits this task to a succession of subproblems.

If subjects are to be able to focus their processing resources on a given subproblem, the hierarchically organized solution plan has to be available somewhere. Although few studies have tried to assess empirically where and how the solution plan is stored, several sources of evidence point to storage in LTM. A rich and organized knowledge structure appears to be a prerequisite for generation of a solution plan. The evidence on skilled memory previously reviewed implies that such a knowledge structure would be highly appropriate for rapid storage and efficient retrieval of information.

Summary

The literature on expert problem-solving contains ample evidence for the operation of a LT-WM. First, in all the areas we have reviewed here--mental calculation, everyday activities, medicine, chess, and knowledge-rich domains--it is apparent that part of being a skilled problemsolver is to possess exceptional memory skills. These memory skills are always highly specific to the domain and can be acquired only through extended practice. They cannot be based on temporary storage in ST-WM because many of the tasks discussed require much greater storage capacity than is available in STM, and because the effects of interference would make storage in STM too unreliable. Furthermore, ample evidence exists (restaurant orders, medical diagnoses, chess) for good LTM after performance of a skilled task.

We have also found a great deal of specific information about the retrieval structures that make expert performance possible in these domains. Experienced mental calculators develop special mnemonic techniques to store the information about a problem and the intermediate results that must be kept available during the course of calculation, whereas intermediate results requiring briefer periods of storage can be distinguished by recency. The retrieval structure of a waiter who memorized dinner orders was studied in some detail and found to be an efficient ad hoc schema, closely adapted to the demands imposed by the particular restaurant menu he had to deal with. The schemas used to organize and remember information in medical diagnosis, various design tasks, and writing appear similar, although the characteristics of LT-WM in those who are expert in these activities have not yet been studied as closely. The retrieval structures in chess, on the other hand, their use by experts, and their acquisition are documented in great detail.

General Summary and Discussion

We have reviewed evidence on working memory and memory performance in a wide range of skilled activities: acquired memory skill in STM tasks such as the digit span, memory in skilled readers, and memory in expert performers in several domains such as mental calculation, medicine and chess. Individuals in all of these areas demonstrate an increased working memory capacity that is restricted to a certain type of information and specific type of activity. Traditionally investigators have focused on a single type of general activity, such as reading, where it would have been plausible that individual differences reflect basic differences in working-memory capacity for that specific domain. However, a comprehensive theory of human performance and memory should provide a uniform and general account of memory phenomena across all types of domains.

A common feature of superior memory performance and increased working memory capacity is that they are restricted to skilled activities. It is generally agreed that in order to attain skilled performance, individuals acquire domain-specific knowledge, procedures, and various perceptual-motor skills. Our central claim is that in addition, they acquire domain-specific skills to expand working memory capacity by developing methods for storing information in LTM in accessible form. We have thus extended Chase and Ericsson's (1982; Ericsson, 1985; Ericsson & Staszewski, 1989) Skilled Memory Theory beyond the acquisition of exceptional memory to account for the increased capacity of working memory in skilled performance.

To meet the particular demands for working memory in a given skilled activity, subjects acquire encoding methods and retrieval structures that allow efficient storage and retrieval from LTM. In the same manner that skilled subjects acquire relevant knowledge of the demands of an activity and develop efficient procedures for completing a task, they also refine methods for encoding information in LTM. The structures that are generated to represent information, guarantee accessibility with respect to specific future retrieval demands. Retrieval demands differ greatly among different activities. Some tasks, like mental abacus calculation, require rapid and frequent updating of digits in a sum, and there is no need to secure extended storage of previous intermediate sums or results. At the other extreme, text comprehension demands extended storage of the constructed representation of the text, where information relevant to the remainder of the text must remain accessible for integration and storage to continue. LT-WM is therefore closely tailored to the demands of a specific activity and is an integrated, inseparable part of the skill in performing the activity.

Our proposal for working memory is superior in several respects to the traditional accounts based only on transient storage in STM. First, our proposal both accounts for the severe constraints on working memory capacity in unfamiliar tasks and explains how working memory can be extended by the acquisition of LT-WM in skilled activities. Second, it goes beyond the description of working memory in normal cognitive activities and explains the ability of subjects skilled in particular activities to cope with interruptions and then successfully resume their activities. Furthermore, the same mechanisms involved in extended working memory also account for memory about an activity once it is completed. Hence, our proposal offers a general description of the function and structure of memory in cognitive activities. A broad range of activities, including skilled activities, and a large set of phenomena and empirical results are subsumed under this framework. We conclude with a brief discussion of these points and some general implications.

Constraints on Working Memory Capacity

Any viable model of working memory has to account for the severe problems of reliably maintaining information in accessible form during cognitive processes. Furthermore, the model must reconcile estimates of working memory capacity with other independent estimates of memory storage capacity from specially designed tests of STM and LTM. In particular, any proposed model needs to be consistent with the large body of research on mechanisms and limits uncovered during a century of study of human memory in the laboratory.

Our theory of working memory is consistent with the traditional theories of human memory (see, Cowan, 1988, and Estes, 1988, for recent reviews) in that it incorporates previously proposed mechanisms and storage types. Our central claim is that under restricted circumstances subjects can extend traditional ST-WM by means of cue-based access of information stored in LTM, the LT-WM.

The mechanisms for storage in and retrieval from LTM rely on generally accepted associative principles of human memory.

In the introduction to this paper we reviewed some arguments based on estimates from laboratory studies that storage in LTM is too slow and unreliable to store information efficiently. We then reviewed an extensive body of results showing that memory experts and other experts can reliably store in LTM information that is relevant to representative activities in their domains of expertise. Concerns were raised about subjects' ability to successfully anticipate future retrieval demands and hence about their ability to select and index information so that they could reliably access it later. In skilled activities and when subjects have had extensive experience with the task demands and acquired stable procedures for completing the task, they can foresee retrieval demands and develop memory skills to index relevant information with retrieval structures.

We described the memory skills of LT-WM for many different activities and supported the claim that attainment of these skills requires many years of practice. Furthermore, the domain-specific memory skills and LT-WM we have considered enable us to understand transient storage in attention and STM in a new light.

Our proposal gives a parsimonious account of findings that have been problematic for the standard account of ST-WM. Many investigators, in particular Broadbent (1975), have argued that G. A. Miller's (1956) assessment of the capacity of STM reflects a maximum (correct performance on 50% of trials on a pure memory task) and that the amount of information that can be reliably stored in STM is much lower, around three or four chunks. Within the context of complex cognitive activities such as problem solving and decision making, the reliable working capacity of ST-WM (measured by a number of independent chunks) is likely to be even lower. Similarly we believe that several studies of subjects' maximal STM capacity in specially designed memory tasks overestimate the reliable capacity of STM during normal task-oriented processing. We also believe that this capacity is at least in part due to special strategies such as active rehearsal, which are not habitually used in complex cognitive processing. Nevertheless, LT-WM makes cognitive processing possible even if the lower estimates of the reliable capacity of ST-WM are true.

Our model is consistent with Baddeley's (1986) proposal for an independent subsystem for rehearsal, that is, the articulatory loop. The evidence for a visual-spatial scratch pad that allows for domain-general storage of spatial information is presently less clear. The reported evidence for storage in the visual-spatial scratch pad and the observed interference from concurrent visual-spatial tasks may reflect acquired domain-specific storage. The research reviewed here on digit-span experts and individuals playing blind-folded chess suggests that LT-WM with retrieval structures based on spatial cues are used.

The mechanisms of LT-WM that we have developed are consistent with the general recency effect discussed by Baddeley (1986; Baddeley & Hitch, 1993) and the implications for working memory of the temporal separation in the Brown-Peterson paradigm discussed by Walter Schneider and Detweiler (1987). Although a more general discussion of current issues in research on STM (Crowder, 1993; Shiffrin, 1993) falls outside the scope of this paper, we believe that our proposal for LT-WM provides concepts and mechanisms that will be relevant to some of the controversies concerning storage in LTM in laboratory tasks designed to study only STM.

Although our model of working memory conforms to all the basic constraints on human information processing, it asserts that subjects can acquire skill in the use of LTM and thereby circumvent the capacity limitations of STM for specific domains and tasks. Our proposal does not abolish constraints on working memory; it merely substitutes new constraints on rapid storage in and efficient retrieval from LT-WM for the old constraints on ST-WM.

The Scope of Relevant Observations for Working Memory

The prevailing conception of working memory as based solely on transient storage in STM is appealing because it is simple. According to this view, a limited number of elements are available in STM during a given state of a cognitive process (see Figure 4). As the cognitive processes unfold, the elements in STM change, but the elements of a given state in STM are sufficient to characterize that state, and investigators have therefore disregarded the subjects' prior processing history. Models based on this prevailing conception typically allow for storage in LTM; but as we noted in the introduction, storage in LTM of the traces of the processes is considered unreliable (described by probabilistic mechanisms). Moreover, storage in LTM is quite limited, at least in many laboratory tasks. Early models (Atkinson & Shiffrin, 1968) had dual and separate representations of elements in STM and LTM, whereas other theories (D. A. Norman, 1968), especially more recently (Shiffrin & Walter Schneider, 1977, Anderson, 1983) propose a more uniform representation, in which elements in STM refer to the activated elements in LTM. In the latter type of model, new associates of an item stored in LTM during cognitive processing should be accessible during subsequent processing.

We have shown that these models of STM are unable to account for subjects' ability to cope with relatively long disruptions of their skilled activities. Furthermore, we have reviewed evidence for accurate and extensive storage of information in LTM that was accessible for controlled retrieval after completion of subjects' tasks. Hence the best evidence for LT-WM comes not from findings about normal processing, but from those for performance under unusual circumstances, such as interruptions imposed by switching between different tasks, by memory testing during processing, and by memory performance after processing has been completed.

One of the benefits of our proposal for LT-WM is that it may account for a much broader range of observations than usual within a single model of concurrent processing. At the same time it necessitates more complicated theories of working memory to describe encoding and storage in LTM with generated associations to relevant retrieval structures. To accurately describe a state in a cognitive process, it will be necessary to specify not only the activated elements in STM but also the generated knowledge structures in LTM. Complete processing models of skilled performance must describe in detail the retrieval cues maintained in ST-WM as well as the generated encodings stored in LTM along with temporal information about when they were stored. Only with such descriptions can investigators fully model the effects of proactive and retroactive interference and the methods subjects acquire to counteract these effects by more extensive and elaborate encodings.

The nature of LT-WM described in our proposal raises issues very different from those studied within the framework of ST-WM. Individual differences in the capacity of working memory are not fundamentally fixed and unchangeable. Instead, they are deliberately acquired. But how? And how can they can be assessed for different domains and tasks? How can instructional procedures be used in remediation? LT-WM for tasks in a given domain of activity is an integrated part of skilled performance. It is clear that our analyses of skilled performance must probe deeply into the organization of knowledge and its encoding and retrieval processes if they are to fully describe the operation of LT-WM. Only if we are willing to fully describe and dissect complex cognitive skills will we ever ascertain the real limits of cognition and create a theoretical framework for working memory that encompasses the full range and complexity of cognitive processes.

References

Aaronson, D., & Ferres, S. (1984). The word-by-word reading paradigm: An experimental and theoretical approach. In D. E. Kieras and M. A. Just (Eds.), New Methods in Reading Comprehension Research (pp. 31-68). Hillsdale, NJ: Erlbaum.

Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence. Journal of Experimental Psychology: Human Perception and Performance, 2, 56-70.

Afflerbach, P. P. (1990). The influence of prior knowledge on expert readers' main idea construction strategies. Reading Research Quarterly, 25, 31-46.

Allard, F., & Starkes, J. L. (1991). Motor-skill experts in sports, dance and other domains. In K. A. Ericsson and J. Smith (Eds.), Toward a general theory of expertise: Prospects and limits (pp. 126-152). Cambridge: Cambridge University Press.

Allport, D. A., 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. (1974). Retrieval of propositional information from long-term memory. Cognitive Psychology, 6, 451-474.

Anderson, J. R. (1983). The architecture of cognition. Cambridge, Mass.: Harvard University Press.

Anderson, J. R. (1987). Methodologies for studying human knowledge Behavioral and Brain Sciences, 10, 467-505

Anderson, J. R. (1990). Cognitive psychology and its implications . San Francisco: Freeman.

Atkinson, R. C., & Shiffrin, R. M. (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.

Atwood, M. E., Masson, M. E., & Polson, P. G. (1980). Further explorations with a process model for water jug problems. Memory & Cognition, 8, 182-192.

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

Baddeley, A. D. (1990). Human memory: Theory and practice . Boston, MA: Allyn and Bacon.

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

Baddeley, A. D., & Hitch, G. J. (1993). The recency effect: Implicit learning with explicit retrieval? Memory & Cognition, 21, 146-155.

Baddeley, A., Papagno, C., & Valla, G. (1988). When long-term learning depends on short-term storage. Journal of Memory and Language, 27, 586-596.

Baddeley, A., & Wilson, B. (1988). Comprehension and working memory: A single-case neuropsychological study. Journal of Memory and Learning, 27, 479-498.

Baggett, P. (1975). Memory for explicit and implicit information in picture stories. Journal of Verbal Learning and Verbal Behavior, 14, 538-548.

Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. Journal of Verbal Learning and Verbal Behavior, 18, 333-356.

Bates, E., Kintsch, W., & Fletcher, C. R. (1980). On the role of pronominalization and ellipsis in text: Some memory experiments. Journal of Experimental Psychology: Human Learning and Memory, 6, 676-691.

Bellezza, F. S. (1982). Updating memory using mnemonic devices. Cognitive Psychology, 14, 301-327.

Bennett, H. L. (1983). Remembering drink orders: The memory skill of cocktail waitresses. Human Learning, 2, 157-169.

Bereiter, C., & Scardamalia, M. (1987). The psychology of written composition. Hillsdale, NJ: Erlbaum.

Binet, A. (1894). Psychologie des grands calculateurs et joueurs d'echecs [The psychology of great calculators and chess players]. Paris: Libraire Hachette.

Bjork, R. A. (1978). The updating of human memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 12, pp. 235-259). New York: Academic Press.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. Cognitive Science, 16, 153-184.

Bower, G. H. (1972). Mental imagery and associative learning. In L. W. Gregg (Ed.), Cognition in learning and memory (pp. 51-88). New York: Wiley

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 3, 193-209.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. Cognitive Psychology, 2, 331-350.

Bransford, J. D., & Johnson, M. K. (1973). Consideration of some problems of comprehension. In W. G. Chase (Ed.) Visual information processing (pp. 383-438). New York: Academic Press.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11(717-726), .

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy and A. Wilkes (Eds.). Studies in long-term memory (pp. 3-18). London: Wiley.

Brown, J. (1958). Some test of the decay theory of immediate memory. Quarterly Journal of Experimental Psychology, 10, 12-21.

Byrne, R. (1977). Planning meals: Problem-solving on a real data-base. Cognition, 5, 287-332.

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual-differences approach. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 1101-1114.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? Intelligence, 15, 229-246.

Carpenter, P. A., & Just. M. A. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), Complex information processing: The impact of Herbert A. Simon (pp. 31-68). Hillsdale, NJ: Erlbaum.

Case, R. (1978). Intellectual development from birth to adulthood: A neo-Piagetian interpretation. In R. Siegler (Ed.), Children's thinking: What develops? Hillsdale, N. J.: Erlbaum.

Charness, N. (1976). Memory for chess positions: Resistance to interference. Journal of Experimental Psychology: Human Learning and Memory, 2, 641-653.

Charness, N. (1979). Components of skill in bridge. Canadian Journal of Psychology, 33, 1-6.

Charness, N. (1981a). Aging and skilled problem solving. Journal of Experimental Psychology: General, 110, 21-38.

Charness, N. (1981b). Search in chess: Age and skill differences. Journal of Experimental Psychology: Human Perception and Performance, 7, 467-476.

Charness, N. (1989). Expertise in chess and bridge. In D. Klahr & K. Kotovsky (Eds.), Complex information processing: The impact of Herbert A. Simon (pp. 183-208). Hillsdale, NJ: Erlbaum.

Charness, N. (1991). Expertise in chess: The balance between knowledge and search. In K. A. Ericsson & J. Smith (Eds.), Toward a general theory of expertise: Prospects and limits (pp. 39-63). New York: Cambridge University Press.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 141-189). Hillsdale, N. J.: Erlbaum.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 16, pp. 1-58). New York: Academic Press.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press.

Chi, M. T. H. (1976). Short-term memory limitations in children: Capacity of processing deficits? Memory & Cognition, 4, 559-572.

Claessen, H. F. A., & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. Medical Education, 19, 61-67.

Cooke, N. J., Atlas, R. S., Lane, D. M., & Berger, R. C. (1993). Role of high-level knowledge in memory for chess positions. American Journal of Psychology, 106, 321-351.

Coughlin, L. D., & Patel, V. L. (1987). Processing of critical information by physicians and medical students. Journal of Medical Education, 62, 818-828.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information-processing system. Psychological Bulletin, 104, 163-191.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.

Crowder, R. G. (1993). Short-term memory: Where do we stand? Memory & Cognition, 21, 142-145.

Crutcher, R. J. (1990). The role of mediation in knowledge acquisition and retention: Learning foreign vocabulary using the keyword method. (Tech Report No. 90-10). Boulder: University of Colorado, Institute of Cognitive Science.

Crutcher, R. J. (1992). The effects of practice on retrieval of foreign vocabulary learned using the keyword method. Unpublished doctoral dissertation. The University of Colorado at Boulder.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Dansereau, D. F. (1969). An information processing model of mental multiplication (Doctoral dissertation, Carnegie-Mellon University, 1969). Dissertation Abstracts International, 30, 1916-B. (University Microfilms No. 69-15746)

Dansereau, D., & Gregg, L. W. (1966). An information processing analysis of mental multiplication. Psychonomic Science, 6, 71-72.

Deakin, J. M., & Allard, F. (1991). Skilled memory in expert figure skaters. Memory & Cognition, 19, 79-86.

de Groot, A. (1978). Thought and choice and chess. The Hague: Mouton. (Original work published 1946).

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. Psychological Bulletin, 89, 63-100.

Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. Journal of Educational Psychology, 80, 465-472.

Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of prose. Journal of Experimental Psychology, 88, 216-222.

Ebbinghaus, H. (1964). Memory: A contribution to experimental psychology (H.A. Ruger & C. E. Bussenius, Trans.). New York: Dover Publications, Inc. (Originally published, 1885).

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

Elo, A. E. (1986). The rating of chessplayers, past and present. 2nd Edition. New York: Arco.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 972-992.

Engle, R. W., & Bukstel, L. H. (1978). Memory processes among bridge players of differing expertise. American Journal of Psychology, 91, 673-689.

Ericsson, K. A. (1985). Memory skill. Canadian Journal of Psychology, 39, 188-231.

Ericsson, K. A. (1988a). Analysis of memory performance in terms of memory skill. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence, (Vol. 5) (pp. 137-179).

Ericsson, K. A. (1988b). Concurrent verbal reports on text comprehension: A review. Text, 8, 295-325.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. American Scientist, 70, 607-615.

Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100(3), 363-406.

Ericsson, K. A., & Oliver, W. (1984). Skilled memory in blindfolded chess. Annual Psychonomic Society Meeting, San Antonio (November).

Ericsson, K. A., & Polson, P. G. (1988a). Memory for restaurant orders. In M. Chi, R. Glaser, & M. Farr (Eds.), The nature of expertise (pp. 23-70). Hillsdale, N. J.: Erlbaum.

Ericsson, K. A., & Polson, P. G. (1988b). An experimental analysis of a memory skill for dinner-orders. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 305-316.

Ericsson, K. A., & Staszewski, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), Complex information processing: The impact of Herbert A. Simon (pp. 235-267). Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data (Revised Edition). Cambridge, MA: MIT Press.

Estes, W. K. (1988). Human learning and memory. In R. C. Atkinson, R. J. Herrnstein, G. Lindsey, & R. D. Luce (Eds.), Stevens' handbook of experimental psychology, Second Edition, Volume 2 (pp. 351-415). New York: Wiley.

Farnham-Diggory, S., & Gregg, L. (1975). Short-term memory function in young readers. Journal of Experimental Child Psychology, 19, 279-298.

Fischer, B. & Glanzer, M. (1986). Short-term storage and the processing of cohesion during reading. The Quarterly Journal of Experimental Psychology, 38A, 431-460.

Fischer, G., Henninger, S., & Redmiles, D. (1991). Cognitive tools for locating and comprehending software objects for reuse. Thirteenth International Conference on Software Engineering, (pp. 318-328). Austin TX: ACM, IEEE.

Fletcher, C. (1981). Short-term memory processes in text comprehension. Journal of Verbal Learning and Verbal Behavior, 20, 264-274.

Fletcher, C. (1986). Strategies for the allocation of short-term memory during comprehension. Journal of Memory and Language, 25, 43-58.

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple narrative texts. Journal of Memory and Language, 27, 235-244.

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. Discourse Processes, 13, 175-190.

Frederiksen, C. H. (1975). Representing logical and semantic structure acquired from discourse. Cognitive Psychology, 7, 371-458.

Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented chess positions. Memory and Cognition, 4, 541-547.

Gernsbacher, M. A. (1990). Language comprehension as structure building. Hillsdale, N. J.: Erlbaum

Gernsbacher, M. A., Varnes, K. R., & Faust, M. E. (1990). Individual differences in general comprehension skills. Journal of experimental Psychology: Learning, Memory, and Cognition, 16, 430-445.

Givón, T. (in press). The grammar of referential coherence as mental processing instructions. In M. Gernsbacher & T. Givón (Ed.) Coherence in natural text. Amsterdam: Benjamin.

Gilhooly, K. J., Wood, M., Kinnear, P. R., & Green, C. (1988). Skill in map reading and memory for maps. Quarterly Journal of Experimental Psychology, 40A, 87-107.

Glanzer, M., Dorfman, D., & Kaplan, B. (1981). Short-term storage in the processing of text. Journal of Verbal Learning and Verbal Behavior, 20, 656-670.

Glanzer, M., Fischer, B., & Dorfman, D. (1984). Short-term storage in reading. Journal of Verbal Learning and Verbal Behavior, 23, 467-486.

Glanzer, M., & Nolan, S. D. (1986). Memory mechanisms in text comprehension. In G. H. Bower (Ed.), The psychology of learning and motivation (pp. 275-317). New York: Academic Press.

Glanzer, M., & Razel, M. (1974). The size of the unit in short-term storage. Journal of Verbal Learning and Verbal Behavior, 13, 114-131.

Groen, G. J., & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), The nature of expertise (pp. 287-310). Hillsdale, NJ: Erlbaum.

Grosz, B. J., & Sidner, C. (1986). Attention, intention, and the structure of discourse. Computational Linguistics, 12, 312-345.

Guindon, R. (1980). Inferences in story comprehension. Unpublished M.Sc. thesis, University of Toronto.

Guindon, R. (1981). Use of verbal reports to study inferences in text comprehension. Paper presented at the Cognitive Science Conference, University of California at Berkeley.

Hanson, V. L., & Bellugi, U. (1982). On the role of sign order and morphological structure in memory for American sign language sentences. Journal of Verbal Learning and Verbal Behavior, 21, 621-633.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. The Psychology of Learning and Motivation, 22, 193-225.

Hatano, G., Amaiwa, S., & Shimizu, K. (1987). Formation of a mental Abacus for computation and its use as a memory device for digits: A developmental study. Developmental Psychology, 23, 832-838.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. Cognition, 15, 95-110.

Hatta, T., Hirose, T., Ikeda, K., & Fukuhara, H. (1989). Digit memory of Soroban experts: Evidence of utilization of mental imagery. Applied Cognitive Psychology, 3, 23-33.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), Cognitive process in writing (pp. 3-30). Hillsdale, NJ: Erlbaum.

Hishitani, S. (1990). Imagery experts: How do expert Abacus operators process imagery? Applied Cognitive Psychology, 4, 33-46.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. Cognitive Psychology, 10, 302-323.

Hjelmquist, E. (1984). Memory for conversations. Discourse Processes, 7, 321-336.

Holding, D. H. (1985). The psychology of chess skill. Hillsdale, NJ: Erlbaum.

Hörmann, H. (1976). Meinen und Verstehen . Frankfurt: Suhrkamp.

Hunter, I. M. L. (1962). An exceptional talent for calculative thinking. British Journal of Psychology, 53, 243-258.

Huttenlocher, J., & Burke, D. (1976). Why does memory span increase with age? Journal of Verbal Learning and Verbal Behavior, 8, 1-31.

Hyde, T. S., & Jenkins, J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. Journal of Verbal Learning and Verbal Behavior, 12, 471-480.

Jarvella, R. J. (1979). Immediate memory and discourse processing. The Psychology of Learning and Motivation, 13, 379-421.

Jeffries, R., Turner, A. A., Polson, P. G., & Atwood, M. E. (1981). The processes involved in designing software. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Jensen, A. R. (1990). Speed of information processing in a calculating prodigy. Intelligence, 14, 259-274.

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard University Press.

Jones, W., & Anderson, J. R. (1987). Short- and long-term memory retrieval: A comparison of the effects of information load and relatedness. Journal of Experimental Psychology: General, 116, 137-153.

Just, M. A., & Carpenter, P. A. (1987). The psychology of reading and language comprehension. Boston: Allyn and Bacon.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension. Psychological Review, 99, 122-149.

Keenan, J. M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversation. Journal of Verbal Learning and Verbal Behavior, 16, 549-560.

Kieras, D. E. (1984). A method for comparing a simulation to reading time data. In D. E. Kieras & M. A. Just (Eds.), Methods in Reading Comprehension Research (pp. 299-325). Hillsdale, NJ: Lawrence Erlbaum.

Kintsch, W. (1974). The representation of meaning in memory . Hillsdale: Erlbaum.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. Psychological Review, 95, 163-182.

Kintsch, W. (1992a). How readers construct situation models for stories: The role of syntactic cues and causal inferences. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Ed.), From learning processes to cognitive processes: Essays in honor of William K. Estes (pp. 261-278). Hillsdale, N.J.: Erlbaum.

Kintsch, W. (1992b). A cognitive architecture for comprehension. In H. L. Pick, Jr., P. van den Broek, & D. Knill (Eds.), Cognition: Conceptual and methodlogical issues (pp. 143-164). Washington, DC: American Psychological Association

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. Discourse Processes, 16, 193-202.

Kintsch, W. (in press-a). Kognitionspsychologische Modelle des Textverstehens: Literarische Texte. In K. Reusser & M. Reusser (Ed.), Verstehen lernen - Verstehen lehren. Bern, Switzerland: Hans Huber.

Kintsch, W. (in press-b). Discourse Processes. In G. d'Ydewalle, P. Eelen, & P. Bertelson (Ed.), Current advances in psychological science: An international perspective. .London, UK: Erlbaum.

Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. Journal of Experimental Psychology: Human Learning and Memory, 3, 150-159.

Kintsch, W., & Keenan, J. M. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. Cognitive Psychology, 5, 257-279.

Kintsch, W., & Kozminsky, E. (1977). Summarizing stories after reading and listening. Journal of Educational Psychology, 69, 491-499.

Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. Psychological Review, 85, 363-394.

Kintsch, W., & Welsch, D. M. (1991). The construction-integration model: A framework for studying memory for text. In W. E. Hockley & S. Lewandowsky (Ed.), Relating theory and data: Essays on human memory in honor of Bennett B. Murdock (pp. 367-385). Hillsdale, NJ.: Erlbaum.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. Journal of Memory and Language, 29, 133-159.

Koltanowski, G. (1985). In the dark. Coraopolis, PA: Chess Enterprises.

Kristofferson, M. (1972). When item recognition and visual search functions are similar. Perception & Psychophysics, 12, 379-384.

Lane, D. M., & Robertson, L. (1979). The generality of levels of processing hypothesis: An application to memory for chess positions. Memory and Cognition, 7, 253-256.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 1118-1133.

Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 71, 331-338.

Luria, A. R. (1968). The mind of a mnemonist. New York: Avon.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. Cognitive Psychology, 24, 56-98.

MacDonald, M. C., & Pearlmutter, N. J. (1993). Probabilistic constraints and working memory capacity in syntactic ambiguity resolution. Presentation at the 34th Annual Psychonomic Society Meeting, Washington, DC (November).

Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), The psychology of learning and motivation (Vol. 1, pp. 327-372). New York: Academic Press.

Martin, R. C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. Memory & Cognition, 21, 176-183.

Masson, M. E. J. (1984). Memory for the surface structure of sentences: Remembering with and without awareness. Journal of Verbal Learning and Verbal Behavior, 23, 579-592.

Masson, M. E. J., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. Journal of Educational Psychology, 75, 314-318.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. Cognitive Psychology, 13, 307-325.

Meyer, B. J. F. (1975). The organization of prose and its effect on memory . Amsterdam: North Holland.

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

Miller, J. R., & Kintsch, W. (1980). Readability and recall for short passages: A theoretical analysis. Journal of Experimental Psychology: Human Learning and Memory, 6, 335-354.

Moeser, S. D. (1979). The role of experimental design in investigations of the fan effect. Journal of Experimental Psychology: Human Learning and Memory, 5, 125-134.

Morris, P. E., Tweedy, M., & Gruneberg, M. M. (1985). Interest, knowledge and the memorization of soccer scores. British Journal of Psychology, 76, 415-425.

Moravcsik, J. E., & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. Canadian Journal of Psychology, 47, 360-374.

Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. Journal of Verbal Learning and Verbal Behavior, 15, 447-458.

Müller, G. E. (1917). 'Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufes: Teil II' [Toward analyses of memory and imagery: Part II]. Zeitschrift für Psychologie, Ergänzungsband 9.

Myers, J. L., O'Brien, E. J., Balota, D. A., & Toyofuku, M. L. (1984). Memory search with interference: The role of integration. Cognitive Psychology, 16, 217-242.

Nathan, M. J., Kintsch, W., & Young, E. (in press). A theory of word algebra problem comprehension and its implications for the design of learning environments. Cognition and Instruction.

Newell, A. (1990). Unified theories of cognition. Cambridge, Mass:, Harvard University Press.

Newell, A., & Simon, H. A. (1972). Human problem solving . Englewood Cliffs, N. J.: Prentice-Hall.

Norman, D. A. (1968). Toward a theory of memory and attention. Psychological Review, 75(6), 522-536.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. Journal of Experimental Psychology: Learning, Memory and Cognition, 15, 1166-1174.

Patel, V. L., & Arocha, J. F. (1993). Construction-integration theory and clinical reasoning. Paper presented at the Annual Meeting of the American Educational Research Association held in Atlanta, GA, April 12-16.

Patel, V. L., & Groen, G. L. (1986). Knowledge based solution strategies in medical reasoning. Cognitive Science, 10, 91-116.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), Toward a general theory of expertise (pp. 93-125). Cambridge, MA: Cambridge University Press.

Patrick, C. (1935). Creative thought in poets. Archives of Psychology, 178.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An informational search and protocol analysis. Organizational Behavior and Human Performance, 16, 366-387.

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

Posner, M. I., & McCandliss, B. D. (1993). Converging methods for investigating lexical access. Psychological Science, 4, 305-309.

Potter, M. C. (1983). Representational buffers: The eye-mind hypothesis in picture perception, reading, and visual search. In K. Rayner (Ed.), Eye movements in reading: Perceptual and language processes (pp. 413-437). New York: Academic Press.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. Journal of Memory and Learning, 29, 633-654.

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 940-953.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. Psychological Review, 95, 385-408.

Recht, D. R., Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. Journal of Educational Psychology, 80, 16-20.

Reder, L. M., & Ross, B. H. (1983). Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 55-72.

Reitman, J. (1976). Skilled perception in go: Deducing memory structures from inter-response times. Cognitive Psychology, 8, 336-356.

Rizzo, N. D. (1939). Studies in visual and auditory memory span with specific reference to reading disability. Journal of Experimental Education, 8, 208-244.

Saarilouma, P. (1989). Chess players' recall of auditorily presented chess positions. European Journal of Cognitive Psychology, 1, 309-320.

Saarilouma, P. (1990). Apperception and restructuring in chess players' problem solving. In K. J. Gilhooly, M. T. G. Keene, & G. Erdos (Eds.), Lines of thought: Reflections on the psychology of thinking(Vol. 2, pp. 41-57). London: Wiley.

Saarilouma, P. (1991a). Visuo-spatial interference and apperception in chess. In R. H. Logie & M. Denis (Eds.), Mental images in human cognition (pp. 83-94). Amsterdam: Elsevier Science Publishers.

Saarilouma, P. (1991b). Aspects of skilled imagery in blindfold chess. Acta Psychologica, 77, 65-89.

Saarilouma, P. (1992). Error in chess: The apperception-restructuring view. Psychological Research, 54, 17-26.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 2, 437-442.

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 3, 552-631.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals, and understanding . Hillsdale, N. J.: Erlbaum.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. Journal of Memory and Language, 25, 279-294.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. Memory & Cognition, 21, 338-351.

Schneider, W. (Walter), & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.) The psychology of learning and motivation. (Vol. 21, pp. 54-119). New York: Academic Press.

Schneider, W. (Wolfgang), Körkel, J., & Wienert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. Journal of Educational Psychology, 81, 306-312.

Shaffer, L. H. (1975). Multiple attention in continuous verbal tasks. In P. M. Rabbitt & S. Dornic (Eds.), Attention and performance (Vol. 5, pp. 157-167). London: Academic Press.

Shiffrin, R. M. (1976). Capacity limitations in information processing, attention and memory. In W. K. Estes (Ed.), Handbook of learning and cognitive processes: Attention and memory (Vol. 4, pp. 177-236). Hillsdale, N. J.: Erlbaum.

Shiffrin, R. M. (1993). Short-term memory: A brief commentary. Memory & Cognition, 21, 193-197.

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

Simon, H. A. (1973). The structure of ill structured problems. Artificial Intelligence, 4, 181-201.

Simon, H. A. (1974). How big is a chunk? Science, 183, 482-488.

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1991) Individual differences in bridging inference processes. Manuscript submitted.

Sloboda, J. (1976). Visual perception of musical notation: Registering pitch symbols in memory. Quarterly Journal of Experimental Psychology, 28, 1-16.

Smith, E. E., Adams, N., & Schorr, D. (1978). Fact retrieval and the paradox of interference. Cognitive Psychology, 10, 438-464.

Smith, S. B. (1983). The great mental calculators. New York: Columbia University Press.

Spillich, G. J. (1983). Life-span components of text processing: Structural and procedural differences. Journal of Verbal Learning and Verbal Behavior, 22, 231-244.

Spillich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain related information for individuals with high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior, 18, 275-290.

Staszewski, J. J. (1988a). The psychological reality of retrieval structures: An investigation of expert knowledge (Doctoral dissertation, Cornell University, 1987). Dissertation Abstracts International, 48, 2126B.

Staszewski, J. J. (1988b). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), The nature of expertise (pp. 71-128). Hillsdale, NJ: Erlbaum.

Staszewski, J. J. (1990). Exceptional memory: The influence of practice and knowledge on development of elaborative encoding strategies. In W. (Wolfgang) Schneider & F. Weinert (Eds.), Interactions among aptitudes, strategies, and knowledge in cognitive performance (pp. 252-285). New York: Springer Verlag.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 57, 421-457.

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. Cognitive Psychology, 16, 145-176.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior, 18, 523-534.

Thorndyke, P. W. & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. Cognitive Psychology, 12, 137-175.

Till, R. E., Mross, E. R., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. Memory & Cognition, 16, 283-298.

Tulving, E. (1979). Relation between encoding specificity and levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing in human memory. Hillsdale, NJ: Erlbaum.

Tulving, E. (1983). Elements of episodic memory. New York: Oxford University Press.

Turner, M. L., & Engle, R. W. (1989) Is working memory capacity task dependent? Journal of Memory and Language, 28, 127-154.

Trabasso, T., & Suh, S. Y. (1993). Using talk-aloud protocols to reveal inferences during comprehension of text. Discourse Processes, 16, 283-298.

Tulving E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 5, 381-191.

van den Broek, P. W. (1990). Causal inferences in narrative comprehension. In A. C. Graesser & G. H. Bower (Eds.), Inferences and text comprehension. The Psychology of Learning and Motivation, 25, 175-196.

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension . New York: Academic Press.

vanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), Foundations of cognitive science (pp. 527-579). Cambridge, MA: Bradford/MIT Press

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. The Psychology of Learning and Motivation, 17, 165-213.

Walker, C. H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. Cognition and Instruction, 4, 25-42.

Watkins, M. J., & Peynircioglu, Z. F. (1983). Three recency effects at the same time. Journal of Verbal Learning & Verbal Behavior, 22, 375-384.

Waugh, N. C. (1985). Acquisition and retention of a verbal habit in early and late adulthood. Bulletin of the Psychonomic Society, 23, 437-439.

Waugh, N., & Norman, D. A. (1965). Primary memory. Psychological Review, 72, 89-104.

Weber, R. J., & Bach, M. (1969). Visual and speech imagery. British Journal of Psychology, 60, 199-202.

Weber, R. J., & Blagowsky, J. (1970). Metered memory search with implicit and explicit scanning. Journal of Experimental Psychology, 84, 343-348.

Weber, R. J., & Castleman, J. (1969). Metered memory search. Psychonomic Science, 16, 311-312.

Weber, R. J., Cross, M., & Carlton, M. (1968). Searching circular sequences. Journal of Experimental Psychology, 78, 588-592.

Wickens, D. D.,, Born, D. G., & Allen, C. K. (1963). Proactive inhibition and Item similarity in short-term memory. Journal of Verbal Learning and Verbal Behavior, 2, 440-445.

Yates, F. A. (1966). The art of memory . Chicago: University of Chicago Press.

Yekovich, F. R., Walker, C. H., Ogle, L. T., & Thompson, M. A. (1990). The influence of domain knowledge on inferencing in low-aptitude individuals. In A. C. Graesser & G. H. Bower (Eds.), Inferences and text comprehension. The Psychology of Learning and Motivation, 25, 175-196.

Yu, B., Zhang, W., Jing, Q, Peng, R., Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese and English language material. Memory & Cognition, 13, 202-207.

Author Notes

This research was supported by grant MH 15872 from the National Institute of Mental Health to Walter Kintsch and partially supported by contract MDA903-90-K-0066 from the Army Research Institute and by the Dr. Edward Conradi Endowment Fund to K. Anders Ericsson. We gratefully acknowledge the thoughtful comments on earlier drafts of this paper by Evelyn Ferstl, Randy Fletcher, Janet Grassia, Reid Hastie, Eileen Kintsch, Danielle McNamara, William Oliver and Robert Rehder.

Footnotes

1. Of which only the first 50-70 ms may be absolutely necessary to obtain the information needed in reading (Rayner, 1993).

2. Examples from Givón (in press).

Figure Captions

Figure 1. The general organization of a retrieval structure with its retrieval cues. Storage of presented information in LT-WM includes associations to particular retrieval cues. These cues can be activated through the retrieval structure and used to access desired information in LT-WM at a later time.

Figure 2. Proposed hierarchical organization of S.F.'s memory encoding of 30 presented digits. The first level contains mnemonic encodings of digit groups, and the second level consists of supergroups in which the relative location of several digit groups are encoded; after Ericsson (1985),

Figure 3. A 25-digit matrix of the type used by Binet (1894) to test his memory experts. He asked subjects to repeat the whole matrix in the various orders shown or to repeat individual rows as five-digit numbers; after Ericsson (1985).

Figure 4. Two different types of encodings of information stored in LT-WM. On the top, a hierarchical organization of retrieval cues associated with units of encoded information. On the bottom, knowledge-based associations relating units of encoded information to each other along with patterns and schemas establishing an integrated memory representation of the presented information in LTM.

Figure 5. The interrelationships between consecutive mental states, long-term memory, and the environment. The large arrows represent the construction process that yield the final, conscious working memory representation. The filled-in circles with the arrows indicate memory buffers involved in the construction process, as further illustrated in Figure 6.

Figure 6. A possible sequence of memory buffers involved in the construction of mental representations in reading.

Figure 7. Contextual priming for associates and topical inferences as a function of the prime-target asynchrony. The context effect is the difference in lexical decision times between context inappropriate and appropriate target words (associates of a word in the text, or topical inferences). After Till, Mross, & Kintsch, 1988.