Psychological Review
1995, Vol. 102, No. 2, 211-245

0033-295X/95/$3.00

Long-Term Working Memory

K. Anders Ericsson
Florida State University

Walter Kintsch
University of Colorado at Boulder

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 coherently integrate 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 or manipulate a chess board. They are able to keep the current locations of all the chess pieces in an accessible form in memory. These working contexts, with their accessible information that changes 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, emphasis added), that is, to information maintained in readily accessible storage for only a short period without rehearsal or reactivation.

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 article we propose that a general account of working memory has to include another mechanism based on skilled use of storage in long-term memory (LTM) that we refer to as long-term working memory (LT-WM) in addition to the temporary storage of information that we refer to as short-term working memory (ST-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 in LTM.

The classic distinction between short-term memory (STM) and LTM (Atkinson & Shiffrin, 1968; Waugh & 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 enables individuals to use LTM as an efficient extension of ST-WM in particular domains and activities after sufficient practice and training. Extending the Chase and Ericsson (1982) skilled memory theory, we show that mechanisms similar to those underlying a 10-fold increase in performance on tests of STM are used by experts and skilled performers to expand 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, individuals rely on specific control processes to encode heeded information in LTM 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

K. Anders Ericsson, Department of Psychology, Florida State University; Walter Kintsch, Institute of Cognitive Science, University of Colorado at Boulder.

This research was supported by Grant MH 15872 from the National Institute of Mental Health and partially supported by Contract MDA903-90-K-0066 from the Army Research Institute and by the Dr. Edward Conradi Endowment Fund.

We gratefully acknowledge the thoughtful comments on earlier drafts of this article by Peter Delaney, Evelyn Ferstl, Randy Fletcher, Janet Grassia, Reid Hastie, Eileen Kintsch, Andreas Lehmann, Rajan Mahadevan, Danielle McNamara, William Oliver, Robert Rehder, Pertti Saariluoma, and Rolf Zwaan.

Correspondence concerning this article should be addressed to K. Anders Ericsson, Department of Psychology, Florida State University, Tallahassee, Florida 32306-1051. Electronic mail may be sent via Internet to ericsson@psy.fsu.edu.

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 that access to them 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 contradicts the assumptions of current models of working memory. We will first trace the development of these models from their origin in general theories of memory based on the vast body of experimental analyses of memory performance. We will also show that LT-WM is consistent with all major assumptions about LTM and STM and that our disagreement concerns a couple of auxiliary assumptions about the fixed speed of storage and retrieval in LTM. In the following section we will summarize the broad range of arguments raised against the possibility of rapid and reliable storage in and retrieval from LTM. Then we show that these objections do not extend to skilled performance and review evidence for rapid and reliable operations of LTM for storage and retrieval. Next we review Chase and Ericsson's (1982) skilled memory theory and its mechanisms of acquired memory skill that enable individuals to attain efficient use of LTM after extended practice. Finally we propose LT-WM by introducing additional mechanisms that allow skilled subjects to overcome effects of proactive and retroactive interference in their use of LTM for storage and retrieval. In the remainder of the article we show that LT-WM and its associated characteristics provide an explanation of working memory during text comprehension as well as of the expanded working memory in expert performance in several domains.

Historical Background to the Development of Current Theories of Working Memory

When Ebbinghaus (1885/1964) introduced the experimental approach to study human memory, he was keenly aware that the most important factors influencing recall and retention in everyday life concern individuals' relevant experience, knowledge, and interests. In fact, his laboratory-based approach was designed to eliminate or, at least, minimize the effects of relevant experience through the study of memory for unfamiliar material, such as nonsense syllables. Furthermore, he relied on rapid sequential presentation of the material to exclude use of acquired skills and strategies. The elimination of the influence of knowledge and skill allowed Ebbinghaus to study the basic mechanisms for strengthening associations and to discover general laws of memory. Since then researchers have concentrated on deriving general laws and capacities for memory from simple tasks explicitly designed for the study of 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 & Norman, 1965), immediate free recall yields items directly retrieved from a temporary STM and items retrieved by retrieval cues from a more durable storage in 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; Peterson & 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 primary bottleneck for retrieval from LTM is the scarcity of retrieval cues that are related by association to the desired item, stored in LTM. Another problem with storage of information in LTM is that subsequent storage of similar information may interfere with the retrieval of the originally stored information. All of these above characteristics of human memory are consistent with our proposal for LT-WM. However, we disagree with two common auxiliary assumptions (Newell & Simon, 1972), namely that retrieval from LTM is slow and takes about 1 s and, in particular, that storage of a new retrievable memory trace is very slow and takes between 5 and 10 s. Although we accept Newell and Simon's (1972) estimate of around 10 s per chunk for memorization of unfamiliar material, we will show that experts require only a fraction of that time for storage in LTM of representative material from their domain of expertise.

The capacities and operational characteristics of STM and LTM discovered in research on memory performance were imposed as constraints on information-processing models of problem solving, decision making, and concept formation. The information required for successful completion of the task has to be kept readily available, which implies both rapid and reliable (essentially error-free) storage of and access to that information. In traditional theories of memory experiments these criteria for availability are met only for information stored in STM. However, the reliable capacity of working memory must be considerably less than the memory span of seven chunks (G. A. Miller, 1956), in which perfect recall is achieved only 50% of the time on average. Hence the capacity of reliable working memory is often 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.[1] For these cognitive activities investigators have shown that storage of relevant information is temporary—consistent with storage only in STM—as subjects' memory of information relevant to the task is poor once the task is completed (Coltheart, 1971; Karat, 1982; Reed, Ernst, & Banerji, 1974, Experiment 2).

As working memory has been considered in a wider range of complex tasks, theorists have found it increasingly difficult, if

[1] An analysis of these tasks (Ericsson & Kintsch, 1991) shows that they were designed to minimize the load on working memory by keeping most of the relevant information perceptually available in external memory (Newell & Simon, 1972).

not impossible, to model the associated cognitive processes with only around four chunks in working memory. In his ACT* theory Anderson (1983) rejected a limit on the number of transiently activated elements in favor of a limit based on the total amount of available activation. In some of his models Anderson found that working memory can sometimes contain over 20 active units at one time. To reconcile such a large capacity of working memory with the much smaller capacity of STM, Anderson argued as follows: The activation of elements decays very rapidly. For this reason the number of elements 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).

Baddeley (1986) has opposed the dominant view of working memory as a single construct with general resources. He has proposed that in addition to the general resources of the central executive there are two slave systems, the articulatory loop and the visuo-spatial sketchpad, in which the central executive can temporarily store information. Much of the evidence for the existence of these two slave systems comes from dual task studies in which a memory task is performed concurrently with the primary task under investigation. Interference with either of these slave systems by a concurrent secondary memory task typically degrades performance on the primary task only slightly. In particular, 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). These findings imply that the central executive has sufficient working memory capacity to complete the processing, leaving working memory for skilled activities virtually unexplained.

In summary, the limited demands on working memory for many unfamiliar tasks used in laboratory studies are mostly consistent with the traditional models that assume a working memory of limited capacity. However, these models do not appear to offer plausible accounts of the increased demand for available information required by skilled processing in more complex tasks. To date models have considered only transient activation as a means to maintain available information. Storage in LTM has been ruled out as impossible for several reasons. In the next section we will briefly review these reasons and show why efficient and reliable storage in and retrieval from LTM is not only possible but also frequently observed in expert performance.

Arguments Against Any General Involvement of LTM in Working Memory and Domain-Specific Exceptions for Skilled Performance

On the basis of a century of laboratory research on memory many theorists have concluded that LTM can meet neither the criteria of speed and reliability for storage nor those for retrieval. We first address the problem concerning storage and then turn to the issues concerning retrieval.

We have already mentioned the excessive duration estimated for the average time of successful storage. Arguments have also been raised against the possibility of attaining reliable storage in LTM. 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 tasks with more meaningful materials subjects' recall of presented information does not improve 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 standard memory tasks. Anderson (1983) even went so far as to argue that subjects' inability to control storage in LTM is beneficial because the subjects cannot predict what information will be useful later. Anderson's comment points to yet another problem, namely that of accurately selecting only that information that might be important during future processing. This is a critical problem for any model of working memory regardless of storage in LTM or STM. In fact, the selection problems should be greater for theories based only on the limited-capacity ST-WM than for theories with additional storage in LT-WM.

Even if subjects could anticipate, select, and reliably store information needed for subsequent processing, the issue of how that information could be selectively and efficiently retrieved from LTM would remain. Most proposed mechanisms have focused on the basic issue of how retrieval of recently stored information in LTM can be distinguished from the vast amount of other information in LTM. Walter Schneider and Detweiler (1987) and Shiffrin (1976) have proposed that recall is mediated by the cues available in the current context. If some of the cues in the recall context were also available in the study context, they might have been part of the memory trace and thus could serve as retrieval cues to access items in the studied list. Once some items are recalled, these items could serve as additional retrieval cues for the remaining items. Given that subjects cannot properly control the available contextual cues, this mechanism can account for the low and varied level of free recall after list learning. However, this mechanism cannot account for the reliable access of specific information needed by subjects at a specific point during subsequent processing. In addition, retrieval from LTM is estimated to be slow compared with STM and thus does not meet the criterion for rapid and efficient retrieval.

These arguments were made in the context of general theories of memory that are invariant across all conceivable tasks. Hence empirical support for them based on performance with unfamiliar tasks rules out the possibility that mediation of storage in LTM for working memory is an invariant feature of human working memory. This conclusion is not controversial, and we disagree only with the stronger claim that the invariant characteristics of LTM rule out an expansion of working memory by storage in LTM in all types of performance.

Our proposal for LT-WM simply argues that subjects can acquire domain-specific memory skills that allow them to acquire LT-WM and thus extend their working memory for a particular activity. To support our argument we will first review evidence showing that rapid and reliable storage in LTM is found for many types of expert performance and is a normal mode of expert processing as reflected by substantial incidental memory of task-related information. In the following section we will show that retrieval from LTM can be both selective and rapid under conditions of cued recall. We will then describe the mechanisms of LT-WM.

Rapid and Reliable Storage in LTM During Skilled Performance

It is generally accepted that memory performance for a particular type of stimulus material depends on the subjects' familiarity with that material and that this performance can be greatly increased by practice (see Ericsson, 1985, for a review). Within general theories of memory such differences have been accounted for in terms of preexisting chunks in LTM. If subjects have already stored a large number of such complex patterns in LTM, then some of these chunks will match the presented items and thus allow the subjects to maintain groups of presented items by activating the matching chunks in STM. Hence with brief presentations of material—with shorter presentation times than those estimated to be necessary for storage of new chunks in LTM—subjects can recall a large number of elements, such as letters, when the elements are part of a small number of recognized chunks, such as words (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 (chunks) of chess pieces in LTM. This vast array of patterns 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 is mediated by familiar meaningful configurations of chess pieces, Chase and Simon (1973) 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); medicine (G. R. Norman, Brooks, & Allen, 1989); 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 can be accounted for solely in terms of STM, in which 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 inconsistent with 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 chess players can recall substantial amounts from up to nine different chess positions that have been presented in sequence at a comparably fast rate of presentation.

The largest individual differences in memory performance are associated with memory experts and professional mnemonists, whose exceptional performance reflects acquired abilities to store specific types of information in LTM (see Ericsson, 1985, 1988a, for a review). Of particular relevance to an increase in the reliable capacity of working memory is the digit-span task, in which subjects have to reproduce perfectly a list of digits presented at a fast rate. After hundreds of hours of practice on this task 2 subjects were able to increase their memory performance from around seven 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 hr 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. Two findings support this claim: Subjects exhibited small decrements in memory performance in response to experimentally induced interference with STM prior to recall of a digit list, and they could very accurately recall all of the presented digit sequences at the end of a test session. Furthermore, improvements in memory performance were restricted to the specific type of practice material.

The findings reviewed above demonstrate that experts and trained subjects are able to store information both rapidly and reliably as evidenced by the vastly superior memory performance on tasks originally designed to test storage in STM. However, the fact that experts are able to store information in LTM during explicit memory tasks does not prove that they regularly rely on such storage to extend their working memory during normal activities.

The most direct method of assessing experts' storage of information in LTM during regular cognitive activities is to make unexpected requests that the experts recall information once they have completed the tasks and related stimuli have been removed from view. 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 in their verbal reports on their thought processes during selection of the move. A subsequent study by Lane and Robertson (1979) showed that chess players' memory of a chess position after the move selection task is as good as it is when the players have been informed in advance about the memory test. 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 tests of 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 et al., 1989).

The pattern of results from experts' incidental recall is consistent with laboratory research on incidental memory (Hyde & Jenkins, 1973) and depth of processing (Craik & Lockhart, 1972). With "deep" and meaningful processing, subjects' memory for the presented stimuli did not benefit from their knowing about a subsequent recall test in advance. Hence memory for the stimuli must be a direct consequence of the cognitive processes mediating the task performance. Within the depth of processing framework, investigators have found that the retrieval operations should match the encoding conditions to be effective (Moscovitch & Craik, 1976). The greater amount of incidental memory observed with experts as compared with naive subjects suggests that experts' stable repertoire of procedures allows them more reliable reinstatement of earlier mental states with their retrieval cues associated to information stored in LTM.

Rapid and Reliable Retrieval of Information From LTM With Cued Recall

Theories of memory make a primary distinction between immediate recall of activated information in STM and the additional step required for accessing other information in LTM. During free recall of lists, subjects typically recall the small number of items in STM immediately, then there is a long pause until additional items in LTM are accessed and reported. In this prototypical memory task, storage in STM overlaps completely with the rapid and reliable accessibility of information, but this overlap does not necessarily occur in skilled activities. In skilled activities with a large amount of accessible information, sequential free recall of all available information does not allow assessment of immediate accessibility. Given that rapid and reliable accessibility of a particular piece of information at a specific time is the functional criterion for working memory, cued recall is the appropriate method for such assessment.

In cued-recall tasks, subjects retrieve an item in response to a retrieval cue provided by the experimenter. Retrieval from STM is often elicited by presentation of a short list of items to the subject, after which the subject is asked 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 10 times slower than the rate assessed for recognition. Several investigators (Weber & Blagowsky, 1970; Weber & Castleman, 1969; Weber, Cross, & Carlton, 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 subvocal 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 et al. (1985) to range from 281–373 ms. Hence, selective retrieval of information stored in STM is far from immediate, and combined search and retrieval might require around 1 s.

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, in general, storage in LTM does not occur immediately for typical laboratory stimuli. Investigators therefore provide subjects with study times sufficient for storage in LTM prior to a recognition test. Sternberg (1969) found that recognition times for highly practiced lists stored in LTM were 200–400 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 typically occurs after a single presentation. In recognition tests of memory for sentences, Anderson (1990) found that accessing sentences stored in LTM takes somewhat longer than accessing 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).

In summary, cued-recall experiments have shown that although retrieval of information from LTM takes longer than access to STM, the difference in latency is small (around 300 ms) when the retrieval cues in STM are closely associated with the target item. If expert performers can form similar associated structures at the time they first store information in LTM, they should be able to access information in LTM with a speed and reliability comparable to that for access from STM.

We next consider the mechanisms that would make LT-WM possible. As our proposal for LT-WM is an extension of Chase and Ericsson's (1982) skilled memory theory, we will describe that theory first.

Skilled Memory Theory: Efficient Storage in and Retrieval From LTM by Retrieval Structures

In their skilled memory theory, Chase and Ericsson (1982) proposed a mechanism to explain how subjects could expand their memory capacity on the digit-span task by over 1,000% after extended training. The proposed mechanism for extension of working memory is attainable only under very restricted circumstances. First, subjects must be able to rapidly store information in LTM; this requires a large body of relevant knowledge and patterns for the particular type of information involved. Our review showed that such abilities are observed not only in memory experts but also in several other types of experts in specific domains. Second, the activity must be very familiar to the experts because only then can they accurately anticipate future demands for retrieval of relevant information. When

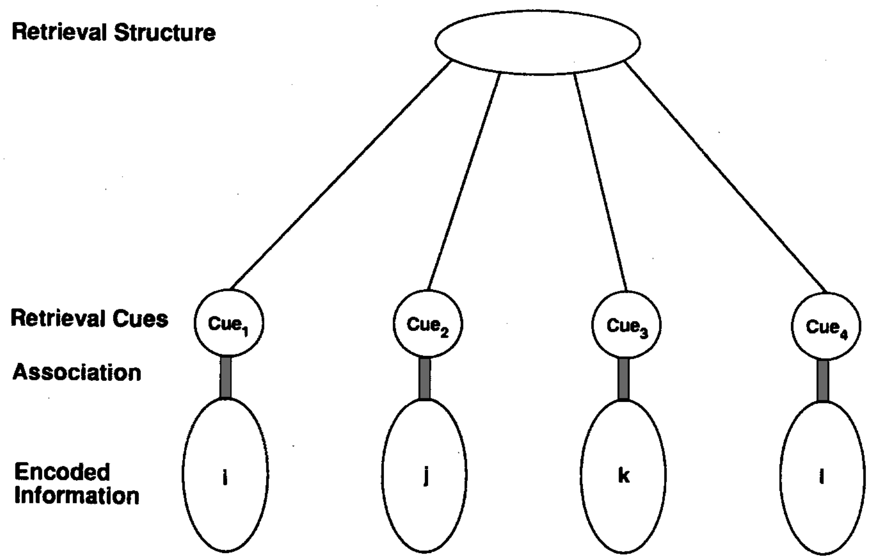

Figure 1. The general organization of a retrieval structure with its retrieval cues. Storage of presented information in long-term working memory (LT-WM) includes associations to particular retrieval cues. These cues can be activated through the retrieval structure and used to access desired information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

these two conditions are met, selective storage of information in LTM is possible. Third, subjects must associate the encoded information with appropriate retrieval cues. This association allows them to activate a particular retrieval cue at a later time and thus partially reinstates 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 a manner consistent with the encoding specificity principle (Tulving, 1983).

The mechanisms based on retrieval structures can be extended to account for expert performance (Ericsson & Staszewski, 1989), and these mechanisms contribute a major component to our proposal for LT-WM. Detailed accounts of the structure of experts' working memory are necessarily very complex because the knowledge and procedures in their domains of expertise are inherently complex. Furthermore, retrieval demands on working memory differ for different domains of expertise, as do the associated retrieval structures. We will first describe the general principle governing LT-WM and explain how it is implemented in comparatively simple cases, such as acquired memory skill for digits. Later we will discuss the complex cases of working memory in text comprehension and in various types of expert performance.

At a very general level we can characterize LT-WM as being 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. 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 encodes that patient's relevant 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 originates from studies of exceptional recall of digits. In a digit-recall task the demands for future retrieval are unusually clear: Subjects are asked to reproduce all the digits in the exact order presented. From 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 (Figure 2). At the lowest level, SF used mnemonic association to running times (3596 → 3 min and 59.6 s or just below a 4-min time for a mile) and other numerical relations to group digits and encode them as units. SF then used spatial relations to encode digit groups into supergroups. 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, trained 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 digit 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 point to the location of a presented digit group. SF was able to perform these retrieval tasks accurately and rapidly. With more

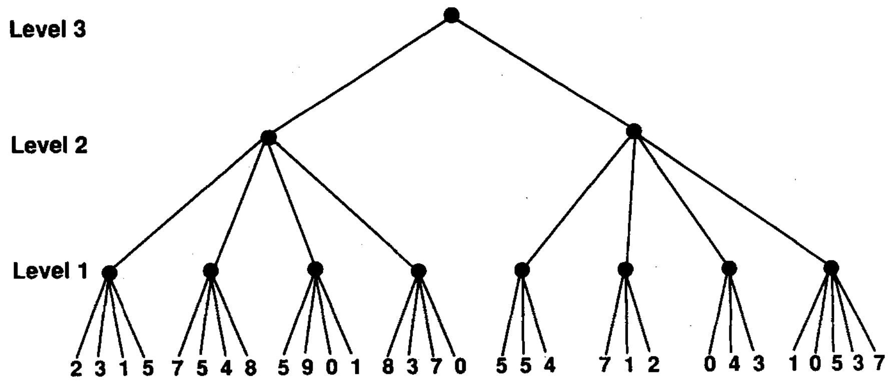

Figure 2. Proposed hierarchical organization of SF's memory encoding of 30 presented digits. The first level contains mnemonic encodings of digit groups, and the second level consists of supergroups in which the relative location of several digit groups are encoded; after Ericsson (1985).

extensive practice and an even higher level of acquired memory skill, a second subject (DD) was able to perform these retrieval tasks even faster, and his retrieval was virtually immediate (Staszewski, 1988a).

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.

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, 1988b) 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 summary, 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, through practice, working memory based on storage in and retrieval from LTM could attain speeds similar to those for STM.

The Chase and Ericsson (1982) proposal for skilled memory has been generally accepted as accounting for exceptional memory (Anderson, 1990; Baddeley, 1990; Carpenter & Just, 1989; Newell, 1990; Walter Schneider & Detweiler, 1987; VanLehn, 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 beforehand" (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 experts' working memory does not rely on LTM, Baddeley (1990) interpreted 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 show that Chase and Ericsson's skilled memory theory can be extended into our conception of LT-WM, which successfully addresses these concerns.

LT-WM: Mechanisms for Overcoming Proactive and Retroactive Interference in Working Memory

In most skilled activities subjects have to repeatedly perform the same task in direct succession. And even within an activity they generate and change intermediate results and products in working memory. If subjects try to retain the most recent result in LT-WM by associating it to the appropriate cue in the retrieval structure, proactive interference from results previously stored with that cue will interfere with its retrieval. This problem has been extensively studied in the laboratory with paired associates, in which 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 interference reflected by these two recall tasks can be dramatically influenced by in-

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

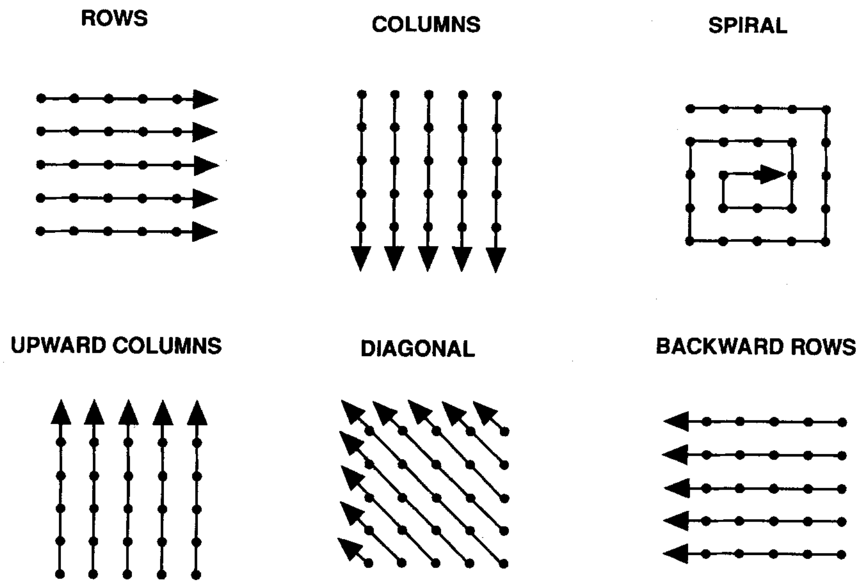

Figure 3. A 25-digit matrix of the type used by Binet (1894) to test his memory experts. He asked subjects to repeat the whole matrix in the various orders shown or to repeat individual rows as five-digit numbers; after Ericsson (1985).

structing unskilled subjects to use different encoding methods (Bjork, 1978). When subjects memorized the paired associates by ordered rehearsal, the typical pattern of interference was obtained: Cued recall of the most recent responses is accurate, but old responses could not be recalled at the final memory test (study by Bjork & 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 that 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 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 this section we will propose that skilled subjects can acquire memory skills suited to their working memory needs that allow them to overcome problems of proactive and retroactive interference by two different mechanisms, that is, recency and elaborative encoding. We will first discuss the possibility of distinguishing the most recently stored result without additional encoding beyond the association to the retrieval cue (cf. Bjork's, 1978, ordered rehearsal). We will then turn to encoding methods that generate additional associations and structures in LTM.

Retrieval of the Most Recent Encoding and Its Temporal Separation

In traditional list learning subjects tend to recall the most recently presented items first, and this recency effect is accounted for by their storage in STM. However, an increasing body of evidence indicates that recency effects can also be observed for information in LTM, in which storage has preceded retrieval by days or months. 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 that 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 previous encodings is sufficient to make this encoding distinctive, retrieval is accurate. We know that even unskilled 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). With increased precision in encoding and storage of items due to acquired skill, recency and temporal separation should be an effective cue mediating reliable retrieval. Evidence from memory experts suggests skilled use of recency information. In the digit-span task memory experts' concurrent verbal reports show that they consistently retrieve the digit encodings in reverse order prior to recalling the list in its presented order (Chase & Ericsson, 1981, 1982). Furthermore, when memory experts encode a digit group during the presentation of the digit list, the experts are frequently reminded of very similar digit groups encountered earlier during the same test session. Typically, the previously encoded digit group accessed by reminding shares the same two initial digits with the current digit group being encoded (Chase & Ericsson, 1982).

The same type of mechanism based on recency can explain a finding from the Brown–Peterson paradigm that has been very difficult to explain with 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) reviewed earlier studies showing 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 min 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 as a result of an experimental change in the category of presented instances (Wickens, Born, & Allen, 1963) and also for multiple recency effects for different categories after presentation of mixed lists (Watkins & Peynircioglu, 1983).

In summary, reliable retrieval is possible even if there are many consecutive associations to the same cue provided that the duration of maintenance is sufficiently short to make the most recently stored result remain temporally distinctive. This result can be achieved by the insertion of sufficiently long intervals between repeated trials or by a controlled schedule of access of relevant intermediate results. In skilled activities that rely on LT-WM based on recency, postsession recall would be poor because the temporal distinctiveness of even the most recent result decreases rapidly over time since original storage. Hence, the poor incidental memory of mental abacus experts is not necessarily inconsistent with LT-WM and storage in LTM. On the other hand, when skilled subjects need to maintain information for longer periods of time, they can generate more elaborate encodings than simple associations. When tested on the Brown–Peterson task memory experts have shown no evidence of decay of memory. Hunt and Love's (1972) subject, VP, recalled around 80% of the presented three-consonant trigrams for all retention intervals. The same level of recall was observed for control subjects on the first trial. With build-up of proactive interference, however, their performance, unlike VP's, decreased and showed the typical decay of memory with increased retention intervals. The mnemonist TE performed perfectly on the Brown–Peterson task for nearly all retention intervals with no evidence for decay of memory (Gordon, Valentine, & Wilding, 1984). TE was highly skilled in the use of mnemonic techniques involving conversion of meaningless consonants into words (CNL → *canal*). Gordon et al. (1984) argued that both VP and TE relied on encoding of the trigrams as words and that the difference in VP's and TE's performance on this task "was due to the latter's practice with this type of technique" (p. 11).

Reliable Retrieval Through Elaborative Encoding

Many skilled activities have demands for working memory that cannot be satisfied by storing the most recent results independently in LT-WM and that require the integration of presented information and sustained access to it as reflected in text comprehension. The mnemonic technique of story generation (see Bower, 1972, for a review) reveals the power and efficiency of comprehension, in which the recall of a list of unrelated items can be improved by embedding the items in a complex interrelated structure generated and stored in LTM. When unskilled subjects use this technique for lists of unrelated items, both storage and retrieval are typically slow; but we know that storage times are dramatically reduced for experts in their domain of expertise. We will first show that with appropriate encoding subjects can retrieve information rapidly even when facts share arguments and a fan effect for recognition for studied facts would be expected. Then we discuss the effects of proactive and retroactive interference for experts in the digit-span task and show how these subjects counteract the interference by the generation of elaborated encodings.

One of the dominant paradigms for studying the effects of interference in LTM involves the 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). Preexperimental 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 show 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 the related construct of a situation model (van Dijk & Kintsch, 1983). It does not appear possible to integrate several 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 on the 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. For example, the retrieval structure for a list of 30 digits is shown in Figure 2. The only difference for a list of 29 digits is that the last digit group would contain only 4 digits. 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 interference has a reliable, yet limited influence.

The influence of interference is more dramatic during postsession 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 three- and four-digit groups during postsession 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 postsession 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 not only recalled the second list almost 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 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 (1982) determined that these subjects encoded 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 four 4-digit groups in the first supergroup (see Figure 2) were 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 of the digit sequences in LTM (see Chase & Ericsson, 1982, and Staszewski, 1990, for a review of the extensive empirical evidence). Figure 4, an augmented version of Figure 1, shows the additional encodings that generate associations forming a new memory structure connected to most of the digits in the list.

In addition to the direct associations between the encoded information and retrieval cues in LTM, subjects build a unique structure in which the elements are directly linked 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, in general, allow constrained recall of information belonging to this specific structure. However, if two semantically similar items have been associated with the same retrieval cue in different lists confusion at recall might still occur. Protection from this type of proactive inference is mediated by reminding memory experts of previously encoded digit groups that are similar to the corresponding digit group in the current list. Especially if a similar digit group had previously been associated with the same retrieval cue, the memory expert would explicitly encode the relation between the two digit groups to allow the retrieval of the most recent one to be reliable (Chase & Ericsson, 1982).

Summary

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 individuals are likely to choose so as to attain reliable and rapid storage and access of information in LT-WM. This encoding method, which is either a retrieval structure or an elaborated memory structure or a combination of the two (as illustrated in Figure 4), determines 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 general storage capacity is greater for specific skilled activities than for unfamiliar ones. A more parsimonious account, however, is that storage in working memory can be increased and is one of many skills individuals attain during the acquisition of skilled performance. Our proposal for LT-WM is consistent with this account. It is also consistent with the conclusion that the superiority of expert memory and of exceptional memory is domain specific and asserts that increased working memory is limited to the specific skilled activity in question. The acquired nature of LT-WM implies that differences exist between tasks and, in addition, that there are potential individual differences in the implementation of LT-WM for a given task. This implication raises new theoretical and methodological challenges for research on working memory, which we will now address.

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 cog-

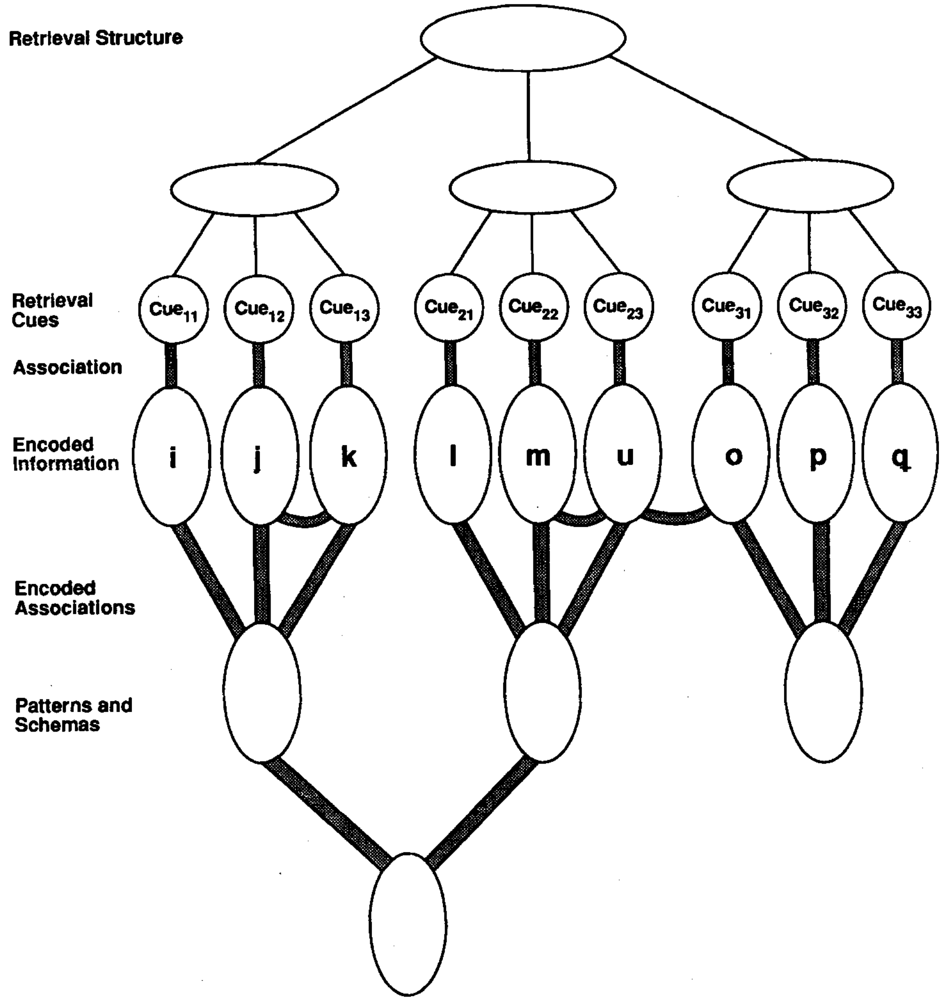

Figure 4. Two different types of encodings of information stored in long-term working memory. On the top, a hierarchical organization of retrieval cues associated with units of encoded information. On the bottom, knowledge-based associations relating units of encoded information to each other along with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

nitive activity. LT-WM is acquired in particular domains to meet specific demands imposed by a given activity on storage and retrieval. 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 require knowledge activation and elaboration processes at various levels. 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, and 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 buff-

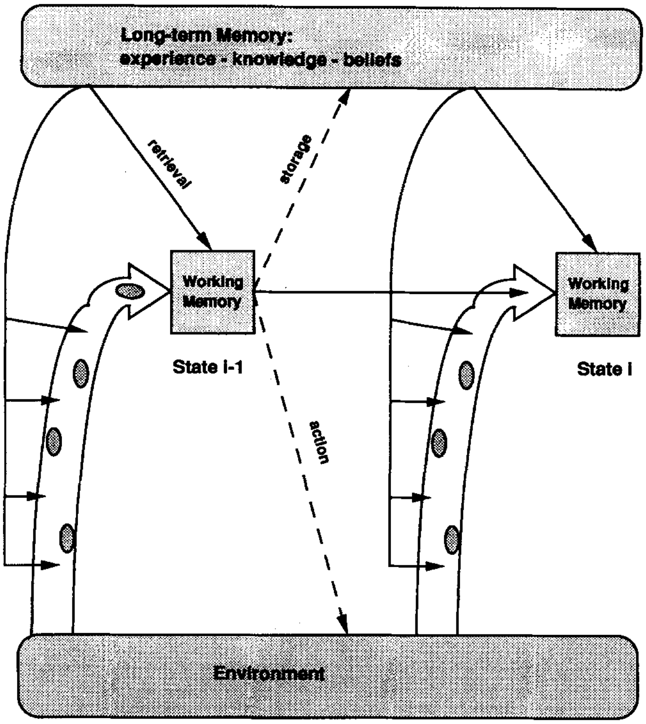

Figure 5. The interrelationships between consecutive mental states, long-term memory, and the environment. The large arrows represent the construction process that yield the final, conscious working memory representation. The filled-in circles with the arrows indicate memory buffers involved in the construction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

ers 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 (a) the state of the environment and its effects on the individual; (b) the individual's knowledge, experience, and beliefs and how they interact with environmental effects; and (c) 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.

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 (1987) 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 can be differentiated, however, if the cognitive processes are interrupted at a given state and the subject's attention is suddenly diverted toward another demanding, unrelated activity for some time until the former cognitive activity can be resumed. If the interruption has a sufficiently long duration, the activated information in ST-WM will be irretrievably lost and the interrupted activity cannot be continued. Information in LT-WM, on the other hand, will only become inaccessible and can be subsequently retrieved. If the significant information is stored in LT-WM, reactivation of the associated retrieval cues in STM will allow subjects to resume their cognitive process after the interruption. Hence induced interruption is an effective experimental technique to differentiate between storage in ST-WM and storage in LT-WM.

Successful experimental intervention 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 sequence 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, and for this reason systematic intervention and disruption of the process are 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. Skilled subjects even scan the words within a phrase in a linear, orderly fashion, 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 from 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, such as chess, 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 a far more common skill than the types of expert performance discussed earlier. In the following review of text comprehension, we first examine the representation of text individuals construct and store in LTM. We then discuss the on-line processes during reading that produce this representation. We describe the processes by which individual text segments are encoded and explain 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 by discussing the construction–integration model of text comprehension to provide a detailed example of the kind of retrieval structures we claim are the basis of LT-WM.

Memory After Completed Text Comprehension

Two characteristics of the long-term memory representation of texts are particularly important for understanding how working memory is used in the generation of these representations. The first is that the long-term memory trace of a text forms some sort of structure (e.g., Gernsbacher, 1990; van Dijk & Kintsch, 1983). Specifically, we represent it as a network of nodes that consists of propositions (for this purpose, imagery must be represented propositionally, too). These propositions are either directly derived from the text or retrieved from the comprehender's LTM. The links between propositions reflect the strength with which propositions had been connected in the text as well as preexisting connections in LTM.

Second, it is useful to distinguish within this representational network nodes and links that were derived from different sources. Some reflect the linguistic surface structure of the text, others derive from the semantic nature and organization of the text, and still others exist because of the comprehender's knowledge about situational relations in the world that are not expressed directly in the text itself (e.g., causal inferences). Accordingly, van Dijk & Kintsch (1983; also Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986) have distinguished among three levels of representation: the linguistic surface structure, the propositional text base, and the situation model. Most researchers differentiate 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; van Dijk & Kintsch, 1983).

There are usually differences in the retention of the surface, text base, 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 text base 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). However, outside of a social context (i.e., in laboratory studies of memory for sentences), memory is in general propositional, and surface features are typically reconstructed (Potter & Lombardi, 1990).

Thus, a multilevel structural representation 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 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 the LTM 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 cannot discuss this transitory period here in any detail. We only stress the related points that it involves a sequence of memory buffers for intermediate processing traces that must be distinguished from working memory, and that this processing takes time. In consequence, different elements of working memory become available at different times. We therefore focus on working memory proper and consider how elements in the structure of a text are integrated across the boundaries of text segments.

Intermediate Representation of Text Segments

How visual information in reading or acoustic information in listening gives rise to conscious meaning in working memory has been studied extensively. In both cases the information undergoes a series of transformations. Intermediate computational results are briefly stored in temporary buffers that often are only accessible in highly restricted ways. Thus, Potter (1983) has described the following sequence of buffers involved in reading: the retinotopic icon, spatial and reatopic visual buffers, a conceptual buffer, an articulatory buffer, and working memory. Baddeley (1986) has described the two slave systems of working memory (the central executive), the articulatory loop and the visuo-spatial sketchpad.

Priming studies (e.g., Gernsbacher, Varner, & Faust, 1990; Swinney, 1979; Till, Mross, & Kintsch, 1988) indicate that it takes about 350 ms for the meaning of a word to be fixated in context. Considerably more time is required, however, to construct a situation model than to disambiguate a word. In the experiment of Till et al., the construction of a situation model amounted essentially to inferring the sentence topic, which required about 1 s of processing time. A similar emergence over time for the encoded structure of sentences has been documented by Gernsbacher (1990; see also Kintsch & Welsch, 1991).

Converging evidence for the time course we propose for the construction of mental representations in reading comes from studies with positron emission tomography (PET) and event-related potential (ERP) methods that are reviewed by Posner and 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, whereas word meanings in a discourse context need approximately 350 ms to stabilize. It has long been recognized in the case of syntactic analysis and text-level processing that although 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.

It is necessary, therefore, to differentiate the function of memory in 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.

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. 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. If, on the other hand, our proposal for LT-WM is correct, then disruption of text comprehension should not cause irreparable harm and can be corrected by retrieval of the necessary structures from LTM because the resumption of reading makes available the appropriate retrieval cues in STM.

Effects of disrupted reading on text comprehension. In a long series of experiments Glanzer and his colleagues (Fischer & Glanzer, 1986; Glanzer, Dorfman, & Kaplan, 1981; Glanzer, 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 with a normal control condition, they found a very consistent general pattern of results across different experiments with different types of interrupting activities. 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.

The ease and speed with which critical elements are retrieved after the interruption should be influenced by the nature of 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.

Compared with continuous reading, interruptions requiring the reader to do short addition problems for about 10 s at the end of each sentence add between 250 and 450 ms to reading times (Glanzer et al., 1981, Experiment 2a). Longer interruptions (about 30 s) and more demanding interruption tasks, such as digit recall, increase reading times by 1,200–1,800 ms (Fischer & Glanzer, 1986, Experiments 2, 3, & 4). If the interruption task involves reading another unrelated sentence (for about 7 s), reading times increase by 314 ms (Glanzer et al., 1984, Experiment 1). The reading time for the intervening unrelated sentences and subsequent memory for these facts were no different from those observed for a control condition. If the duration of the unrelated reading is increased to around 30 s, the increase in reading time for the primary text was 355 ms per sentence, with no comprehension deficits (Glanzer et al., 1984, Experiment 3).

In summary, the research by Glanzer and his colleagues (Fischer & Glanzer, 1986; Glanzer et al., 1981, 1984) 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, in which 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 needed for accessing the retrieval cues that allow continued reading is around 350 ms, which is roughly equivalent to other estimates of retrieval from LTM studies 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 (Fischer & Glanzer, 1986; Glanzer & Nolan, 1986).

Hypothesized role of storage of surface structure in STM. 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, 1984) clearly formulated this idea in their early work. They proposed that what is carried over in STM 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 they 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 (e.g., 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 several 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 Experiment 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. found that many of the sentences subjects first encountered after the interruption did not have references to the preceding sentences 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 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 of their four experiments, Fischer and Glanzer estimated additive increases in resumed reading times to be 408 ms when the increase was due to unavailability of general theme and 402 ms when the increase was due to unavailable linkage information for 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, Experiment 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. Recently, Lorch (1993) has replicated and extended these findings showing that topic information serves as a context for encoding and access of subordinated information in the text.

The interruption procedure, therefore, does little more than slow down the reading process somewhat, 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 LTM that is readily accessible in LT-WM.

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.

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.

In several studies by Baddeley (1986) 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 were 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 with the performance of subjects who did 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 to learn, 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 (1986) 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.

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 with 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. This pattern of verbalizing primarily the presented text during fluent comprehension is in agreement with the comprehension model of Kintsch (1988). Only if comprehension breaks down and the normal flow of reading is interrupted by problem-solving processes to repair a failure to achieve a well-integrated mental representation are subjects able to verbalize their deliberate efforts to attain understanding. To gain more information about the stable end products of comprehension researchers have modified the reading task in think-aloud studies by asking 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 to which we will turn next.

LT-WM during reading. The evidence for the accessibility of the representation of prior text is largely indirect, because successful comprehension would be impossible without such access. However, a few studies have measured accessibility during text comprehension in a more direct fashion.

Several 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 lead 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 think-aloud protocols on the access and generation of verbally reported 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 inferences verbalized after each sentence. Trabasso and Suh successfully predicted the frequencies with which specific inferences were verbalized. They were also 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 & Nolan, 1986) conducted two experiments in which they interrupted subjects 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, Nolan and Glanzer (cited in Glanzer & Nolan, 1986) studied the maintained accessibility of information presented in earlier parts of the text. Recognition of topic and detail statements in the previous paragraph was compared with the same performance for the two types of statements in the current 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 with 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, consistent with the theory of LT-WM.

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.

Consistent with Baddeley (1986) 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, 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 that they needed to comprehend 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 was the subject's reading span. College students managed 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 = .50–.60) and with the ability to answer content questions about a previously read text (r = .70–.90).

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 .80 and .90. 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.

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) argued that working memory reflects a general capacity to hold information in LTM activated during processing. Just and Carpenter (1992) held a similar view with the exception that the capacity limitation is specific to language information. According to our proposal superior text comprehension reflects superior skill in encoding information in LTM, thus allowing a larger amount of information to remain accessible by means of cues 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 Miller (1983) found that, in cued recall, other words in the sentences were equally as good as the final words as predictors of reading comprehension. 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 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 they were 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% 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 working memory, respectively, memorize unrelated (Experiment 1) and thematically related (Experiment 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-working memory group than it was for the high-working memory group. Most interestingly the relation between verbal ability and working memory capacity disappeared when the effects of the slope of the fan were controlled. For the thematically related sentences the low-working memory group still revealed a typical fan effect, whereas the high-working memory group showed a negative fan (i.e., 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-working memory group appears to encode the sentences in isolation or in small groups of thematically related ones, whereas the high-working memory group is able to attain a more integrated representation of the sentences, especially for the thematically related sentences.

Text comprehension is closely related to verbal ability, which is often measured by tests of vocabulary and language use (grammar). Tests of vocabulary and word meanings are correlated with text comprehension as well as with 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 & Miller, 1983). Singer, Andrusiak, Reisdorf, and Black (1992) 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 text base because these inferences are necessary to make the text base 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 (i.e., 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 most compelling 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 or 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. These findings show that individuals' working memory capacity is not constant but varies systematically for texts about different topics.

Yekovich, Walker, Ogle, and Thompson (1990) compared two groups of students of low verbal ability who differed in their knowledge about football (high and 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 were 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. 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.

Thus, the reading span results are better accounted for by LT-WM theory than by capacity theory. LT-WM theory also provides a parsimonious explanation for the other experimental evidence that Just and Carpenter (1992) have adduced in favor of their capacity theory of comprehension. 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 for a good reader because it enables that reader to store more information in working memory during reading. In another version, which fits the data reviewed by Just and Carpenter 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) reported 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 MacDonald and Pearlmutter (1993), who reexamined the finding by MacDonald, Just, and Carpenter (1992) that high-span subjects take longer to read temporarily ambiguous sentences than low-span subjects. In contradiction to Just and Carpenter's account in terms of capacity theory, MacDonald and Pearlmutter showed 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), and 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.

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, 1994a, 1994b; 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 that 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 hand-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 a fully interpreted, contextually integrated text representation. As the reader proceeds to the next sentence, some of the text elements being processed are 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 (given that operations of any kind can be performed only on material held in working memory). If the episodic text memory 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, and 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 representations that are generated in comprehension thus become the retrieval structures of LT-WM.

These structures, situation models, can be quite complex, consisting partly of the propositional network derived from the text (the text base) and partly of associated knowledge. Moreover, 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.

How LT-WM functions in text comprehension may be clarified by referring back to Figure 1. When reading a sentence as part of a larger text, the reader holds in the focus of attention a set of propositions derived from the words and phrases of the sentence. (Only a subset of all propositions that could be constructed are ordinarily produced in any given comprehension episode). These propositions are interrelated in a hierarchical network, including macropropositions, corresponding to the semantic and rhetorical relations established by the text. In Figure 1, the propositions in the focus of attention correspond to the retrieval cues, except that their interrelationships are usually more complex. Some of these propositions are linked to propositions derived from the text on previous processing cycles and now stored in LTM. These elements of LTM make up LT-WM, corresponding in Figure 1 to the row labeled *encoded information.* They are in turn linked to other long-term memory elements, which thus can be retrieved indirectly. However, the important point here is that the links between propositions currently in the focus of attention and propositions in the long-term episodic text memory, which are established incidentally by the very nature of the comprehension process, make available to the reader a large subset of the text memory in LTM, thus generating what we call LT-WM.

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*** *. . .*).[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 when 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, in press, 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, then 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 text bases 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, chap. 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 for a 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

[2] The examples are from Givón (in press).

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 (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,* nor *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 propositions 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 min, 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 inclusion of associated knowledge in the episodic text memory transforms what would otherwise merely be a text base (a structure directly derived from the text) into a (more or less elaborated) situation model and corresponds to the difference between Figure 1 and Figure 4 in our earlier discussion of retrieval structures. What gets added are the patterns and schemas linked to the text base in Figure 4: associated knowledge readers have that serves to organize what they read, enriching the retrieval structure they have created. The region of LTM that is thus directly accessible as LT-WM thereby is enlarged and structured.

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 text base. 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 STM 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 & Bloom, 1988; Kintsch, 1992b.)

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 (in which the contents would be lost during interruptions, as in the Glanzer [Fischer & Glanzer, 1986; Glanzer et al., 1981, 1984] experiments), 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 structure 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 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.

Expanded Working Memory in Expert Performance and LT-WM: A Broader View

The central claim in our proposal for LT-WM is that individuals can acquire memory skill to accommodate expanded demands for working memory in a specific task domain. In the introduction we described the memory skills of subjects with exceptional performance on the digit-span task. The brief sequential presentation of digits and the requirement of perfect serial recall might account for the hierarchical organization of the retrieval structures acquired by all exceptional performers on this task. However, the specific details of a retrieval structure differ clearly among all these subjects (Ericsson, 1988a). Similarly, the demands of comprehension of a text require a successive integration of serially presented sentences into a generated new memory structure. In this section we will attempt to further extend our analysis to other tasks with different retrieval demands.

We first describe all the task domains in which experts with superior working memory capacity have been identified and in which the structure of their working memory has been studied. In the same section we will report evidence relevant to our claim that extended working memory capacity is acquired over an extended period in response to relevant training. The following section considers the working memory demands for five types of task domains and reviews evidence on the particular structure of LT-WM acquired to meet those demands.

Superior Working Memory Capacity in a Task Domain and Its Acquisition

Our criteria for demonstrations of superior working memory capacity with several studies of its structure are only met in five different task domains. For each of these task domains we will identify the corresponding well-defined task for which superior performance has been observed. Then we will discuss evidence on the associated superior memory performance followed by evidence on the acquisition of expert performance in the domain.

In Japan addition of large numbers is often done by updating a running sum on an abacus. Highly skilled individuals can rapidly add numbers relying on a mental 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 material is within the normal range (Hatano &

Osawa, 1983). 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 three or four digits. The rule of thumb is that expanding the mental abacus by one additional digit takes about 1 year of deliberate effort (Hatano & Osawa, 1983). The acquired mental abacus must be similar to an actual abacus as the use of the mental abacus results in a similar pattern of errors of calculation. Furthermore, when experts were asked to monitor for a specific configuration of beads on the abacus during a calculation, their pattern of reaction times was similar for both mental calculations and calculations with an actual abacus (Stigler, 1984).

Mental calculators can rapidly multiply two large numbers in their head (mental multiplication). These individuals have been found to exhibit 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, 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 hr of specific practice (Ericsson, 1985). 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) have accumulated a vast body of knowledge about mathematics and methods of calculation and have 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.

Many experienced waiters and waitresses memorize dinner orders because by doing so they can be more attentive to the customers and thereby earn a larger tip. Superior memory for dinner orders (Ericsson & Polson, 1988a, 1988b) in experienced waiters and waitresses has been studied with laboratory analogs of the real situation in restaurants. Ericsson and Polson identified a waiter (JC) whose memory performance was vastly superior to that of control subjects as well as the performance of a group of waiters and waitresses with experience of memorizing dinner orders. However, JC's memory ability was limited to situations similar to the dinner-order memorization task and in a far-transfer condition JC's performance was reduced toward the level of other waiters, waitresses, and control subjects. Consistent with subjects in other studies of memory skill, JC reported that he had only gradually acquired his exceptional memory performance. He started with parties of small numbers of customers and then moved on to increasingly larger parties of up to 20 customers over a period of several years.

The domain-specific superiority of memory performance of experts in medicine and chess has already been documented in the introduction so we focus on identifying well-defined tasks that capture the expert performance in each of these domains (Ericsson & Smith, 1991). Based on the memory demands for the identified tasks we will then consider how the expanded working memory might have been acquired.

The distinctive criterion of medical experts is their superior accuracy in diagnosing medical cases. In a laboratory analog of the medical diagnosis task, subjects are presented with a text describing a particular patient, and diagnostic performance on this task is closely related to medical expertise (Patel & Groen, 1991; Schmidt & Boshuizen, 1993). The task of medical diagnosis presents challenges to working memory in that symptoms and relevant medical facts have to be maintained in accessible form until the correct diagnosis is identified. Regular engagement in the diagnostic activity would offer opportunity and motivation for improvement in working memory capacity and thus could account for its improvement as a function of the extended experience and increased knowledge of medical specialists compared with medical students, interns, and residents.

The best laboratory task for capturing chess skill involves the selection of the next move for an unfamiliar chess position (de Groot, 1946/1978; Ericsson & Smith, 1991). Performance on the move-selection task is highly correlated with official chess ratings (Charness, 1991; de Groot, 1946/1978; Saariluoma, 1990). For chess players there is no similar sequence of formal training as there is in medicine. Given that the length of experience within a domain is generally a weak predictor of performance, recent research has tried to identify the most effective training activities for improving performance in that domain (Ericsson, Krampe, & Tesch-Römer, 1993). According to biographies and interviews with elite chess players, the best practice activity that chess players can engage in by themselves for extended periods is the study of published chess games between chess masters. During the study of such a game the chess player would try to predict each move made by the chess masters. In case of a failure to predict a move, the chess player would then study the associated chess position more carefully and plan out move sequences to a greater depth to uncover the reasons for the chess master's actual move. A recent, as yet unpublished, study by Charness, Mayr, and Krampe (1994) found a high correlation between the estimated amount of this type of chess study and the chess rating of a large group of tournament players. Engaging in this type of study should improve ability to remember chess positions and to improve the ability to mentally manipulate a chess position to explore deeper plans; both abilities have been found to increase as a function of higher chess ratings (Charness, 1989, 1991).

In summary, superior working memory capacity reflects a domain-specific memory skill acquired to meet specific demands on working memory. This line of argument can also account for the lack of superior memory performance of experts in some other domains (Ericsson & Pennington, 1993). We will now turn to a detailed examination of LT-WM in different domains.

Structure of LT-WM in a Task as a Function of the Demands on Working Memory

The retrieval demands differ greatly for different types of tasks. At one extreme mental abacus calculators update the cumulative sum frequently and rapidly. They only need to maintain the sum and the number that is currently being added. In mental multiplication the calculator has to remember the original problem and some of the intermediate products for some time along with the transient memory support for the calculation of subproblems. The expert waiter has to maintain the dinner orders in memory while the party of customers makes their orders and until he can write down the orders for the chefs in the kitchen. However, at a later time he needs to recall sufficient information to be able to place the salads and entrees in front of the correct customer. The memory demands for experts in medicine and chess are also large and extend over considerable time. In medical diagnosis the experts sequentially encounter facts about the patient. The facts have to be maintained in accessible form without the generation of premature and possibly incorrect inferences. Once the diagnosis is generated the expert needs to be able to access all the relevant medical facts to show that the hypothesized diagnosis explains these facts better than any alternative diagnosis. Similarly, during the generation of potential next moves for a chess position a chess expert needs a memory representation that accurately reflects the resulting chess position after a sequence of moves has been made mentally. If the planning relies too heavily on interpreted configuration the chess expert will not be able to discover unanticipated consequences and might overlook the best available move.

The evidence on the structure of working memory in these task domains is diverse and was mostly collected without concern for the relevant distinctions of LT-WM. Thus we will attempt to organize our review primarily around the two types of mechanisms of LT-WM, namely the generation of new structures in LTM and the use of retrieval structures. The ability of subjects to distinguish and recall information from different episodes or trials as reflected by incidental memory and postsession recall would imply encodings of new structures in LTM that are resistant to proactive and retroactive interference. Evidence for retrieval structures would come from studies of cued recall and of rapid and flexible encoding and access of information. We will also report results on the effects of concurrent memory tasks when they are available.

Mental Abacus Calculation

The requirement for frequent and rapid updating of the cumulative sum during addition would suggest that mental abacus experts have acquired a retrieval structure that allows them to access each location representing a digit on the physical abacus with a unique retrieval cue with a structure similar to that in Figure 1. This structure would allow the expert to independently update the digit associated with any location. These simple encodings could be rapidly made, but previous values would be lost due to retroactive interference caused by the frequent reuse of the same retrieval cue.

Postsession recall and incidental memory. Consistent with the aforementioned hypothesis on the retrieval structure, Hatano and Osawa (1983) found very poor incidental memory for digit sequences encountered by mental abacus experts. In their experiment, 2 elite subjects were asked to maintain a 10-digit sequence for 30 s for 10 consecutive trials, after which they were given a surprise recall and recognition test. One of the subjects could accurately recall the last 10-digit sequence, but otherwise these subjects exhibited no evidence for memory of the presented lists.

Retrieval structures. Several studies have found evidence supporting the acquisition of a retrieval structure as a function of increased level of skill in mental abacus calculation. The accessibility of presented digits in memory has been found to change as a function of skill level from a sequential organization in novices to a more flexible representation that allows order-independent access in experts. 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).

Interference from concurrent memory tasks. 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 visuo-spatial task was found to be reliably greater for elite than for less skilled subjects in one study (Hatano & Osawa, 1983), but not in another (Hatano et al., 1987).

In summary, the proposed account of expert mental abacus calculation by an acquired retrieval structure and simple associations between the maintained digits and their spatial locations provide a good description of the available data.

Mental Multiplication

In addition to performing calculations, like abacus experts, a mental calculator has to maintain information about the original multiplication problem and several intermediate products. Hence we focus on the methods of encoding and storing this type of information that have to be kept accessible for considerable time without attention.

Postsession recall and incidental memory. For highly trained students postsession recall of the numbers in presented multiplication problems was found by Staszewski (1988b) to be moderate to high, with around 55% of the number correctly recalled. He observed that recall for the most recently presented problems was higher, indicating retroactive interference for recall of problems presented early in the test session. In support of this hypothesis Staszewski found around 90% accuracy in a postsession test involving recognition of the presented problems for his trained subjects as well as for a professional mental calculator. More indirect evidence for selective storage of information in LTM during skilled mental calculation comes from Dansereau's (1969) process analyses of three skilled mental calculators. On the basis of concurrent verbal reports of their solutions and performance in designed experiments, he was able to build simulation models for each of his subjects' mental calculation performance. With estimated parameters of their storage and retrieval processes he was able to accurately predict their observed performance. As a function of the level of mental multiplication skill he found consistent differences in the estimated speed of storage, especially in LTM. An analysis of the concurrent verbal reports provided a plausible account of these differences. The 2 best subjects encoded problems and intermediate products by noticing patterns and relations ($9218 \rightarrow 9 \times 2 = 18$), whereas the slowest subject relied on rote rehearsal.

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 three-, four-, and five-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 it shows that an expert deliberately selects different encoding methods to fit the specific retrieval demands of different intermediate results within a given calculational procedure.

Retrieval structures. The primary evidence on retrieval structures in mental multiplication comes from Staszewski's (1988b) training study. In an extended training study Staszewski 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. The distinct encodings differed systematically as a function of number of digits of a problem, because the number and size of intermediate results would differ and thus so would the retrieval context. Staszewski evaluated this hypothesis about the specific organization of retrieval cues for a given problem type by evaluating transfer of performance to multiplication problems with different structure than the problems used in training. Throughout training his subjects had solved problems in which a two-digit, three-digit, four-digit, or five-digit number is multiplied by a two-digit number and only two intermediate products are generated. With these problems 1 of his subjects matched the solution times of an experienced mental calculator. When tested on problems for which a three-digit number is multiplied by another three-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 organization of retrieval cues to accommodate efficient and reliable storage and retrieval of one additional intermediate result.

In summary, studies of mental calculators reveal the use of deliberate methods to encode information in LTM with distinctive cues that facilitate efficient and accurate retrieval.

Expert Memory for Dinner Orders

The expert waiter (JC) needs to maintain the dinner orders from previous customers while the remaining customers give their orders. While the dinner orders are given JC must also be able to access and update customers' orders in case they change their mind, which implies a flexible encoding of the items of a dinner order associated with the customer and his or her location at the table. JC's memory representation must also overcome the similarity and confusability of dinner orders by different customers, because that is the major problem reported by other waitresses and waiters who have attempted to memorize dinner orders for parties of more than two or three customers.

Postsession recall and incidental memory. Postsession recall of the expert waiter (JC) was recorded both after a normal work shift at the restaurant and after test sessions with the laboratory analog of the dinner orders task (Ericsson & Polson, 1988a, 1988b). When JC was unexpectedly tested on his memory for dinner orders taken from customers in the restaurant where he worked, he recalled most of the information with high accuracy. His postsession 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. The discrepancy in postsession recall between the restaurant situation and the laboratory analog is probably due to a couple of factors. In the restaurant tables are spatially distributed, and interactions with real customers offer richer cues for encoding and retrieval. Furthermore, in the laboratory analog JC was instructed to minimize his study times, and he reported that these study times were considerably shorter than the times occupied by real customers who had to select their dinner orders.

Retrieval structures. 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 and developed a method to overcome the problems of interference between similar dinner orders by different customers. 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 locations 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, so that he could 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 clockwise order. Consequently, JC would recall the studied dinner orders by category with the items reported in a consistent clockwise direction. Hence, Ericsson and Polson inferred that JC had a retrieval structure consisting of a sequence of spatial location associated with each customer for each category of items.

To test this hypothesis Ericsson and Polson (1988a, 1988b) changed the order of presentation. Instead of presenting the dinner orders of customers according to their seating in clockwise fashion, the location of the next customer was determined randomly. With random presentation JC's study times and recall accuracy were indistinguishable from the standard presentation format for tables with three and five customers, and only slightly longer study times were observed for eight customers due to a larger memory load in the random condition (see Ericsson & Polson, 1988b, for a detailed account). Most importantly JC recalled the dinner orders by category in the standard clockwise order even in the random presentation condition. Think-aloud protocols confirmed that JC recoded the dinner orders in the random condition into their appropriate spatial locations in the retrieval structure. In two other experiments Ericsson and Polson 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.

Interference from concurrent memory tasks. In one experiment JC had to concurrently count from 1 to 10 while memorizing dinner orders. Accuracy of recall was not influenced, and only a slight increase in study times was observed (Ericsson & Polson, 1988a). However, JC's ability to overcome interference due to unrelated activities is best evidenced by his unimpaired performance in the restaurant while interacting with customers.

In summary, JC's memory skill was found to be mediated by retrieval structures and associated storage in LTM aided by mnemonic encodings and discovered patterns of related items of dinner orders by customers seated next to each other.

Medical Expertise

The primary problem of working memory in medical diagnosis concerns the need to store individual facts in accessible form prior to the recognition of the correct diagnosis. 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.

Retrieval structures. Empirical evidence on diagnostic expertise is consistent with the acquisition of a retrieval structure that allows experts to encode basic medical facts about a patient into higher level diagnostic facts (Patel & Arocha, 1993) so that the correct diagnostic category and specific diagnosis can be accessed. The acquisition of such a retrieval structure as a function of expertise leads to two empirically testable predictions. First, the quality and structure of recalled information about a patient should differ as a function of the level of expertise. Second, experts should be relatively insensitive to the order in which information is presented and at the time of recall reveal a consistent order of recall that reflects the associated retrieval structure (as JC did with dinner orders).

Medical experts are able to identify and recall important information better than novices (see Groen & Patel, 1988, for a review). Furthermore, Schmidt and Boshuizen (1993) were able to show that experts' free recall became more abstract and summarylike as their level of expertise increased.[3] 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).

Analyses of medical experts' 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 that 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 presentation formats for 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.

In summary, retrieval structures of medical experts are most clearly distinguished by their ability to encode higher level information that remains invariant across diagnostic alternatives. It is reasonable to assume that a critical function of LT-WM in

---

[3] When subjects' recall of medical information is tested in an explicit memory task with ample study time studies show an inverse-U function relating 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 result is inconsistent with the earlier reported results for recall of briefly presented information and incidental recall in which recall increases uniformly as a function of expertise. Changes in the encoding processes of experts can account for this discrepancy. The increased selectivity and frequency of abstract encoding of medical information by experts leads to a lower level of recall than intermediates when the number of all recalled pieces of presented information is counted. There also appear to be additional factors related to the special skills of expert medical diagnosis, because when students are asked to memorize information, their recall of all presented information increases compared with 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).

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

The critical demand on working memory in skilled chess playing occurs during the selection of the next move while planning 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. Charness (1981b) found a reliable relation between the maximum number of chess moves planned ahead (depth of search) and chess skill. Saariluoma (1991b) 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. Increases in chess skill beyond this level are associated with a more sophisticated focus of evaluation and abstract planning. 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.

Retrieval structures. Ericsson and Oliver (1984; Ericsson & Staszewski, 1989) proposed that the retrieval structures for chess position in memory correspond to an actual chess board that 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 are able to encode and store the locations of individual chess pieces of a chess position in the absence of meaningful configurations of chess pieces. Second, when skilled chess players have memorized a chess position they can rapidly access the contents of any of the 64 squares of the chess board. Third, chess masters are able to mentally manipulate and update their memory representation of a chess position and even play chess games blindfolded.

In traditional memory tests for a chess position subjects are simultaneously shown all the pieces in their respective locations on the chess board. It is, however, possible to convey the same information by listing all the pieces on the chess board with their respective locations—black knight on d4, white pawn on e6, and so on. Theories of chess skill based on visual recognition of meaningful configurations of chess pieces should predict that chess players would not be able to recall a complete chess position if the locations of all of the individual pieces were described one at a time in random order. If, on the other hand, chess experts had a retrieval structure corresponding to a mental chess board, they could store each piece at a time at the appropriate location within the retrieval structure. After the end of the presentation the experts would be able to perfectly recall the entire position if the presentation rate had been slow enough. This outcome was found by Saariluoma (1989, Experiments 1 and 2), who observed a close relation between chess skill and recall on this task, in which chess masters exhibited nearly perfect recall. Saariluoma also found some evidence suggesting that patterns of adjacent chess pieces aided storage. 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. Most intriguingly, with this sequential format of presentation Saariluoma found that skilled chess players were even able to encode and recall lists of randomly located chess pieces and that the level of recall was closely related to chess skill.[4] The ability to store random chess positions provides particularly strong evidence for the ability to encode individual chess pieces into the retrieval structure.

Other findings by Saariluoma (1989) show that meaningful relations between chess pieces are also encoded 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 chess positions in sequence, either four regular or four random, and then asked to recall them. These players could still recall the regular game positions well—chess masters' accuracy was around 60%—whereas recall of the random positions was below 10%. This finding suggests that meaningful chess positions can be integrated into new distinct structures in LTM whereas random positions cannot.

The proposed 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 chess board. Ericsson and Oliver (1984; Ericsson & Staszewski, 1989) conducted a series of studies of cued recall with a chess master as a subject. The chess master first memorized a chess position in around 10 s. During the cued-recall phase a randomly determined location on the chess board was presented visually, and the chess master reported the name of the piece in that location or said "nothing" if that location was not occupied. The chess master's responses were fast (around 1 s) and very accurate. These retrieval times were close to those in another condition in which the chess master had the tested chess position available in view. In other experiments with memorized chess positions the same cues for location were presented but the chess master's task differed. For example, the chess master was asked to report the number of black pieces that attacked the location specified by the cue. He responded on average in around 4 s and was very accurate, which implies a rapid and highly selective search.

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 reportedly 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;

[4] This result might appear inconsistent with the finding that randomly arranged chess positions are poorly recalled by chess masters and novices alike. However, those studies used a total presentation time of 5 s. The presentation rate in Saariluoma's (1989) experiments was much slower—either 2 s or 4 s per presented chess piece. To present all the pieces of a normal position with this method would require 1–2 min.

Ericsson & Staszewski, 1989) presented the moves in an unfamiliar chess game without a perceptually available chess board. After over 40 chess pieces had been moved, the chess master's representation of the resulting chess position was tested by cued recall with the earlier described procedure. His responses matched virtually perfectly the chess position at that point of the chess game. His brief reaction times reflected rapid access to all the locations on the board.

In a series of studies with chess players at several different levels of skill, Saariluoma (1991a) 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, whereas chess experts' recall deteriorated with increased numbers of moves to around 40%, and novices were totally unable to perform the task. Saariluoma (1991a, Experiment 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. The most accomplished blindfold chess player in the world is George Koltanowski (1985), who several times played blindfold chess against 30 or more opponents, and he won most of the games with the rest of the games resulting in a draw. Playing blindfold chess against a large number of opponents requires the acquisition of additional specialized skill (Koltanowski, 1985).

Interference due to concurrent memory tasks. Planning during move selection imposes a major load on working memory. Saariluoma (1991b) has studied the effect on move generation from concurrent tasks that interfere with the articulatory loop and the visuo-spatial sketchpad (Baddeley, 1986). He found no effect from articulatory suppression but did find a reliable effect from concurrent visuo-spatial tasks (Saariluoma, 1991b). The effect did not differ for chess masters and novices, however, and the visuo-spatial tasks may draw on central resources of perception and attention rather than selectively interfere with working memory. In a couple of other experiments, Saariluoma (1991a) 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 maintained access to the updated chess positions.

In summary, research on planning and memory of chess positions offers some of the most compelling evidence for LT-WM.

Summary

All of the five task domains revealed an increased working memory capacity for experts. The increase was limited to activities within the specific domain, and the amount of increase was related to the level of attained skill and to the amount of relevant prior practice. From an informal analysis of the specific demands on working memory for each task domain, necessary retrieval characteristics were identified, and hypotheses about the organization of retrieval structures were proposed separately for each task.

In all five task domains we find clear evidence for storage in LTM mediated by retrieval structures. The best example of a storage mediated only by a retrieval structure, like the one illustrated in Figure 1, was found in mental abacus calculation, in which frequent reuse of the same retrieval cues led to very low incidental memory at a later memory test. The strongest evidence for retrieval structures concerns the ability of experts to independently store pieces of information when they are presented out of their normal context in scrambled order. After such a presentation, experts in medicine and chess and a waiter (JC) were able to recall all of the information in an order reflecting its typical meaningful organization in the retrieval structure.

Storage in LTM of new structures was predicted and observed in all task domains except mental abacus calculation. In mental calculation and the dinner order task systematic and distinctive encoding of information in LTM was used to avoid interference and ensure reliable and efficient retrieval. Postsession recall of clinical cases by medical experts and of chess positions by chess masters revealed new complex structures in LTM. The best example of the combination of new distinct structures in LTM and of retrieval structures as illustrated in Figure 4 is found for blindfold chess, in which a grand master was able to play out 10 different chess games at the same time.

Indirect evidence for our proposal of LT-WM comes from the virtual lack of interference from concurrent memory tasks on the working memory of experts. In three of the five task domains, relevant studies have not found any effects of interference with the articulatory loop on the accuracy of expert performance. Consistent effects of interference with the visuo-spatial sketchpad were found in chess, but the results suggest that the associated secondary tasks do not interfere with working memory, but rather with the central resources of attention and perception.

In summary, LT-WM offers a general account for the acquisition and structure of expanded working memory capacity in expert performance in the five task domains studied. A review of working memory in additional expert activities (Ericsson & Kintsch, 1994), such as design, problem solving, and text composition, found evidence consistent with LT-WM, although much less firm evidence (especially experimental evidence) is available in these domains.

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, in which 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 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 must acquire encoding methods and retrieval structures that allow efficient storage and retrieval from LTM. In the same manner that skilled subjects must 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 processing. 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 article 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 several 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 proposed 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 the overestimate of capacity of ST-WM 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 visuo-spatial sketchpad that allows for domain-general storage of spatial information is presently less clear. The reported evidence for storage in the visuo-spatial sketchpad and the observed interference from concurrent visuo-spatial tasks may reflect acquired domain-specific storage.

The research reviewed here on digit-span experts and individuals playing blindfold chess suggests that LT-WM with retrieval structures based on spatial cues is 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 article, 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.

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 of the traces of the processes in LTM 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 recent ones (Anderson, 1983; Shiffrin & Schneider, 1977), propose a more uniform representation, in which elements in STM correspond 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 is common 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 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 dissect complex cognitive skills and fully describe them 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 & 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 & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 126–152). Cambridge, England: 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, MA: 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 & 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: Allyn & 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., & Wilson, B. (1988). Comprehension and working memory: A single-case neuropsychological study. *Journal of Memory and Learning, 27,* 479–498.

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.

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.

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. (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 & 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.

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?* (pp. 37–72). Hillsdale, NJ: 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,* 1–16.

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). Cambridge, England: Cambridge University Press.

Charness, N., Mayr, U., & Krampe, R. Th. (1994, Summer). The importance of coaching in entrepreneurial skill domains: An international comparison of life-span chess skill acquisition. In *The acquisition of expert performance: Conference abstracts.* (Available from Anders Ericsson, Department of Psychology, Florida State University, Tallahassee, FL 32306-1051).

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: 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* (pp. 215–281). 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.

Coltheart, V. (1971). Memory for stimuli and memory for hypotheses in concept identification. *Journal of Experimental Psychology, 89,* 102–108.

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.

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. *Dissertation Abstracts International, 30,* 1916B. (University Microfilms No. 69-15746)

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, Netherlands: 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 psy-*

chology (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885)

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

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

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.

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. 4, pp. 137–179). Hillsdale, NJ: Erlbaum.

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., & Kintsch, W. (1991). *Memory in comprehension and problem solving: A long-term working memory* (Institute of Cognitive Science Tech. Rep. No. 91-13). Boulder: University of Colorado.

Ericsson, K. A., & Kintsch, W. (1994). *Long-term working memory* (Institute of Cognitive Science Tech. Rep. No. 94-01). Boulder: University of Colorado.

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

Ericsson, K. A., & Oliver, W. (1984, November). *Skilled memory in blindfolded chess.* Paper presented at the annual meeting of the Psychonomic Society, San Antonio, TX.

Ericsson, K. A., & Pennington, N. (1993). Experts and expertise. In G. Davis & R. Logie (Eds.), *Memory in everyday life* (pp. 241–272). Amsterdam: North-Holland.

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

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

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

Ericsson, K. A., & Smith, J. (1991). Prospects and limits in the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 1–38). Cambridge, England: Cambridge University Press.

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.

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* (2nd ed., Vol. 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. *Quarterly Journal of Experimental Psychology, 38A,* 431–460.

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

Fletcher, C. R. (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 & Cognition, 4,* 541–547.

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

Gernsbacher, M. A., Varner, 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 (Eds.), *Coherence in natural text.* Amsterdam: Benjamin.

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* (Vol. 20, 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.

Gordon, P., Valentine, E., & Wilding, J. (1984). One man's memory: A study of a mnemonist. *British Journal of Psychology, 75,* 1–14.

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 master's 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.

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.

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

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

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

Hunt, E., & Love, T. (1972). How good can memory be? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 237–260). New York: Holt.

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. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13, pp. 379–421). New York: Academic Press.

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 & Bacon.

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

Karat, J. A. (1982). Model of problem solving with incomplete constraint knowledge. *Cognitive Psychology, 14,* 538–559.

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: Erlbaum.

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

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

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

Kintsch, W. (1992b). 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 (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (pp. 261–278). Hillsdale, NJ: Erlbaum.

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

Kintsch, W. (1994a). Discourse processes. In G. d'Ydewalle, P. Eelen, & P. Bertelson (Eds.), *Current advances in psychological science: An international perspective: Vol. 2. The state of the art* (pp. 135–155). London: Erlbaum.

Kintsch, W. (1994b). Text comprehension, memory, and learning. *American Psychologist, 49,* 294–303.

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., & 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 (Eds.), *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.

Lane, D. M., & Robertson, L. (1979). The generality of levels of processing hypothesis: An application to memory for chess positions. *Memory & 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.

Lorch, R. F. (1993). Integration of topic and subordinate information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1071–1081.

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

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, November). *Probabilistic constraints and working memory capacity in syntactic ambiguity resolution.* Paper presented at the 34th annual meeting of the Psychonomic Society, Washington, DC.

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.

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 2]. *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.

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

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

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75,* 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, April). *Construction–integration theory and clinical reasoning.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

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: Prospects and limits* (pp. 93–125). Cambridge, England: Cambridge University Press.

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.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology, 6,* 436–450.

Reitman, J. (1976). Skilled perception in go: Deducing memory structures from interresponse 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.

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

Saariluoma, 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.

Saariluoma, P. (1991a). Aspects of skilled imagery in blindfold chess. *Acta Psychologica, 77,* 65–89.

Saariluoma, P. (1991b). 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.

Saariluoma, 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, NJ: 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., & 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., Körkel, J., & Weinert, 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, NJ: Erlbaum.

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

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

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). *Individual differences in bridging inference processes. Memory & Cognition, 20,* 539–548.

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.

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

Spilich, 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. *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. 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.

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.
Trabasso, T., & Suh, S. Y. (1993). Using talk-aloud protocols to reveal inferences during comprehension of text. *Discourse Processes, 16,* 3–34.
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.
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.
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 and Verbal Behavior, 22,* 375–384.
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.), *The psychology of learning and motivation* (Vol. 25, pp. 175–196). New York: Academic Press.
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.

Received January 3, 1994
Revision received August 10, 1994
Accepted September 26, 1994 ■

Correction to MacDonald et al. (1994)

The title was printed incorrectly for the article by Maryellen C. MacDonald, Neal J. Pearlmutter, and Mark S. Seidenberg (*Psychological Review,* 1994, Vol. 101, No. 4, pp. 676–703). The correct title is "The Lexical Nature of Syntactic Ambiguity Resolution."