NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH Public Access
Author Manuscript
*Prog Brain Res*. Author manuscript; available in PMC 2009 March 18.

Published in final edited form as:
*Prog Brain Res*. 2008 ; 169: 323–338. doi:10.1016/S0079-6123(07)00020-9.

# What are the differences between long-term, short-term, and working memory?

**Nelson Cowan***
*Department of Psychological Sciences, University of Missouri, 18 McAlester Hall, Columbia, MO 65211, USA*

## Abstract

In the recent literature there has been considerable confusion about the three types of memory: long-term, short-term, and working memory. This chapter strives to reduce that confusion and makes up-to-date assessments of these types of memory. Long- and short-term memory could differ in two fundamental ways, with only short-term memory demonstrating (1) temporal decay and (2) chunk capacity limits. Both properties of short-term memory are still controversial but the current literature is rather encouraging regarding the existence of both decay and capacity limits. Working memory has been conceived and defined in three different, slightly discrepant ways: as short-term memory applied to cognitive tasks, as a multi-component system that holds and manipulates information in short-term memory, and as the use of attention to manage short-term memory. Regardless of the definition, there are some measures of memory in the short term that seem routine and do not correlate well with cognitive aptitudes and other measures (those usually identified with the term “working memory”) that seem more attention demanding and do correlate well with these aptitudes. The evidence is evaluated and placed within a theoretical framework depicted in Fig. 1.

### Keywords

attention; capacity of working memory; control of attention; decay of short-term memory; focus of attention; long-term memory; short-term memory; working memory

## Historical roots of a basic scientific question

How many phases of a memory are there? In a naïve view of memory, it could be made all of one cloth. Some people have a good ability to capture facts and events in memory, whereas others have less such ability. Yet, long before there were true psychological laboratories, a more careful observation must have shown that there are separable aspects of memory. An elderly teacher might be seen relating old lessons as vividly as he ever did, and yet it might be evident that his ability to capture the names of new students, or to recall which student made what comment in an ongoing conversation, has diminished over the years.

The scientific study of memory is usually traced back to Hermann Ebbinghaus (1885/1913 translation), who examined his own acquisition and forgetting of new information in the form of series of nonsense syllables tested at various periods upto 31 days. Among many important observations, Ebbinghaus noticed that he often had a “first fleeting grasp … of the series in moments of special concentration” (p. 33) but that this immediate memory did not ensure that the series had been memorized in a way that would allow its recall later on. Stable memorization sometimes required further repetitions of the series. Soon afterward, James (1890) proposed a

*Corresponding author. Tel.: +1 573-882-4232; Fax: +1 573-882-7710; E-mail: CowanN@missouri.edu.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

distinction between primary memory, the small amount of information held as the trailing edge of the conscious present, and secondary memory, the vast body of knowledge stored over a lifetime. The primary memory of James is like the first fleeting grasp of Ebbinghaus.

The Industrial Revolution made some new demands on what James (1890) called primary memory. In the 1850s, telegraph operators had to remember and interpret rapid series of dots and dashes conveyed acoustically. In 1876, the telephone was invented. Three years later, operators in Lowell, Massachusetts started using telephone numbers for more than 200 subscribers so that substitute operators could be more easily trained if the town's four regular operators succumbed to a raging measles epidemic. This use of telephone numbers, complemented by a word prefix, of course spread. (The author's telephone number in 1957 was Whitehall 2–6742; the number is still assigned, albeit as a seven-digit number.) Even before the book by Ebbinghaus, Nipher (1878) reported on the serial position curve obtained among the digits in logarithms that he tried to recall. The nonsense syllables that Ebbinghaus had invented as a tool can be seen to have acquired more ecological validity in an industrial age with expanding information demands, perhaps highlighting the practical importance of primary memory in daily life. Primary memory seems taxed as one is asked to keep in mind aspects of an unfamiliar situation, such as names, places, things, and ideas that one has not encountered before.

Yet, the subjective experience of a difference between primary and secondary memory does not automatically guarantee that these types of memory separately contribute to the science of remembering. Researchers from a different perspective have long hoped that they could write a single equation, or a single set of principles at least, that would capture all of memory, from the very immediate to the very long-term. McGeoch (1932) illustrated that forgetting over time was not simply a matter of an inevitable decay of memory but rather of interference during the retention interval; one could find situations in which memory improved, rather than diminish, over time. From this perspective, one might view what appeared to be forgetting from primary memory as the profound effect of interference from other items on memory for any one item, with interference effects continuing forever but not totally destroying a given memory. This perspective has been maintained and developed over the years by a steady line of researchers believing in the unity of memory, including, among others, Melton (1963), Bjork and Whitten (1974), Wickelgren (1974), Crowder (1982, 1993), Glenberg and Swanson (1986), Brown et al. (2000), Nairne (2002), Neath and Surprenant (2003), and Lewandowsky et al. (2004).

## Description of three kinds of memory

In this chapter I will assess the strength of evidence for three types of memory: long-term memory, short-term memory, and working memory. *Long-term memory* is a vast store of knowledge and a record of prior events, and it exists according to all theoretical views; it would be difficult to deny that each normal person has at his or her command a rich, although not flawless or complete, set of long-term memories.

*Short-term memory* is related to the primary memory of James (1890) and is a term that Broadbent (1958) and Atkinson and Shiffrin (1968) used in slightly different ways. Like Atkinson and Shiffrin, I take it to reflect faculties of the human mind that can hold a limited amount of information in a very accessible state temporarily. One difference between the term "short-term memory" and the term "primary memory" is that the latter might be considered to be more restricted. It is possible that not every temporarily accessible idea is, or even was, in conscious awareness. For example, by this conception, if you are speaking to a person with a foreign accent and inadvertently alter your speech to match the foreign speaker's accent, you are influenced by what was until that point an unconscious (and therefore uncontrollable) aspect of your short-term memory. One might relate short-term memory to a pattern of neural firing

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

that represents a particular idea and one might consider the idea to be in short-term memory only when the firing pattern, or cell assembly, is active (Hebb, 1949). The individual might or might not be aware of the idea during that period of activation.

*Working memory* is not completely distinct from short-term memory. It is a term that was used by Miller et al. (1960) to refer to memory as it is used to plan and carry out behavior. One relies on working memory to retain the partial results while solving an arithmetic problem without paper, to combine the premises in a lengthy rhetorical argument, or to bake a cake without making the unfortunate mistake of adding the same ingredient twice. (Your working memory would have been more heavily taxed while reading the previous sentence if I had saved the phrase "one relies on working memory" until the end of the sentence, which I did in within my first draft of that sentence; working memory thus affects good writing.) The term "working memory" became much more dominant in the field after Baddeley and Hitch (1974) demonstrated that a single module could not account for all kinds of temporary memory. Their thinking led to an influential model (Baddeley, 1986) in which verbal-phonological and visual-spatial representations were held separately, and were managed and manipulated with the help of attention-related processes, termed the central executive. In the 1974 paper, this central executive possibly had its own memory that crossed domains of representation. By 1986, this general memory had been eliminated from the model, but it was added back again by Baddeley (2000) in the form of an *episodic buffer*. That seemed necessary to explain short-term memory of features that did not match the other stores (particularly semantic information in memory) and to explain cross-domain associations in working memory, such as the retention of links between names and faces. Because of the work of Baddeley et al. (1975), working memory is generally viewed as the combination of multiple components working together. Some even include in that bundle the heavy contribution of long-term memory, which reduces the working memory load by organizing and grouping information in working memory into a smaller number of units (Miller, 1956; Ericsson and Kintsch, 1995). For example, the letter series IRSCIAFBI can be remembered much more easily as a series of acronyms for three federal agencies of the United States of America: the Internal Revenue Service (IRS), the Central Intelligence Agency (CIA), and the Federal Bureau of Investigation (FBI). However, that factor was not emphasized in the well-known model of Baddeley (1986).

What is clear from my definition is that working memory includes short-term memory and other processing mechanisms that help to make use of short-term memory. This definition is different from the one used by some other researchers (e.g., Engle, 2002), who would like to reserve the term working memory to refer only to the attention-related aspects of short-term memory. This, however, is not so much a debate about substance, but rather a slightly confusing discrepancy in the usage of terms.

One reason to pursue the term working memory is that measures of working memory have been found to correlate with intellectual aptitudes (and especially fluid intelligence) better than measures of short-term memory and, in fact, possibly better than measures of any other particular psychological process (e.g., Daneman and Carpenter, 1980; Kyllonen and Christal, 1990; Daneman and Merikle, 1996; Engle et al., 1999; Conway et al., 2005). It has been thought that this reflects the use of measures that incorporate not only storage but also processing, the notion being that both storage and processing have to be engaged concurrently to assess working memory capacity in a way that is related to cognitive aptitude. More recently, Engle et al. (1999) introduced the notion that aptitudes and working memory both depend on the ability to control attention, or to apply the control of attention to the management of both primary and secondary memory (Unsworth and Engle, 2007). However, more research is needed on exactly what we learn from the high correlation between working memory and intellectual aptitudes, and this issue will be discussed further after the more basic issue of the short-term versus the long-term memory distinction is addressed.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Meanwhile, it may be helpful to summarize a theoretical framework (Cowan, 1988, 1995, 1999, 2001, 2005) based on past research. This framework, illustrated in Fig. 1, helps to account for the relation between long-term, short-term, and working memory mechanisms and explains what I see as the relation between them. In this framework, short-term memory is derived from a temporarily activated subset of information in long-term memory. This activated subset may decay as a function of time unless it is refreshed, although the evidence for decay is still tentative at best. A subset of the activated information is the focus of attention, which appears to be limited in chunk capacity (how many separate items can be included at once). New associations between activated elements can form the focus of attention. Now the evidence related to this modeling framework will be discussed.

## The short-term memory/long-term memory distinction

If there is a difference between short- and long-term memory stores, there are two possible ways in which these stores may differ: in *duration*, and in *capacity*. A duration difference means that items in short-term storage decay from this sort of storage as a function of time. A capacity difference means that there is a limit in how many items short-term storage can hold. If there is only a limit in capacity, a number of items smaller than the capacity limit could remain in short-term storage until they are replaced by other items. Both types of limit are controversial. Therefore, in order to assess the usefulness of the short-term storage concept, duration and capacity limits will be assessed in turn.

## Duration limits

The concept of short-term memory limited by decay over time was present even at the beginning of cognitive psychology, for example in the work of Broadbent (1958). If decay were the only principle affecting performance in an immediate memory experiment, it would perhaps be easy to detect this decay. However, even in Broadbent's work contaminating variables were recognized. To assess decay one must take into account, or overcome, contaminating effects of rehearsal, long-term retrieval, and temporal distinctiveness, which will be discussed one at a time in conjunction with evidence for and against decay.

### Overcoming contamination from rehearsal

According to various researchers there is a process whereby one imagines how the words on the list are pronounced without saying them aloud, a process called covert verbal rehearsal. With practice, this process comes to occur with a minimum of attention. Guttentag (1984) used a secondary task to show that rehearsal of a list to be recalled was effortful in young children, but not in adults. If, in a particular experimental procedure, no loss of short-term memory is observed, one can attribute that response pattern to rehearsal. Therefore, steps have been taken to eliminate rehearsal through a process termed articulatory suppression, in which a simple utterance such as the word " the" is repeatedly pronounced by the participant during part or all of the short-term memory task (e.g., Baddeley et al., 1975). There is still the possible objection that whatever utterance is used to suppress rehearsal unfortunately causes interference, which could be the true reason for memory loss over time instead of decay.

That problem of interference would appear moot in light of the findings of Lewandowsky et al. (2004). They presented lists of letters to be recalled and varied how long the participant was supposed to take to recall each item in the list. In some conditions, they added articulatory suppression to prevent rehearsal. Despite that suppression, they observed no difference in performance with the time between items in the response varying between 400 and 1600 ms (or between conditions in which the word "super" was pronounced one, two, or three times between consecutive items in the response). They found no evidence of memory decay.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

A limitation of this finding, though, is that covert verbal rehearsal may not be the only type of rehearsal that participants can use. Perhaps there are types that are not prevented by articulatory suppression. In particular, Cowan (1992) suggested that the process of mentally attending to words or searching through the list, an attention-demanding process, could serve to reactivate items to be recalled in a manner similar to covert verbal rehearsal. The key difference is that it would not be expected that articulatory suppression would prevent that type of rehearsal. Instead, to prevent that type of rehearsal an attention-demanding task would have to be used.

Barrouillet et al. (2004, 2007) have results that do seem to suggest that there is another, more attention-demanding type of rehearsal. They have interposed materials between items to be recalled that require choices; they can be numbers to read aloud or multi-choice reaction times. It is found that these interfere with retention to an extent commensurate to the proportion of the inter-item interval used up attending to the distracting items. As the rate of the distracting items goes up, fewer of the to-be-recalled items are recalled. The notion is that when the distracting task does not require attention, the freed-up attention allows an attention- based rehearsal of the items to be recalled. When the interposed task is more automatic and does not require as much attention (e.g., an articulatory suppression task) there is much less effect of the rate of these interposed items.

Based on this logic, one could imagine a version of Lewandowsky’s task in which not articulatory suppression but attention-demanding verbal stimuli are placed between items in the response, and in which the duration of this filled time between items in the response varies from trial to trial. The verbal, attention-demanding stimuli should prevent both attention-based rehearsal and articulation-based rehearsal. If there is decay, then performance should decline across serial positions more severely when longer filled intervals are placed between items in the response. Unfortunately, though, such results might be accounted for alternatively as the result of interference from the distracting stimuli, without the need to invoke decay.

What seems to be needed, then, is a procedure to prevent both articulation-based and attention-based rehearsal without introducing interference. Cowan and Aubuchon (in press) tried out one type of procedure that may accomplish this. They presented lists of seven printed digits in which the time between items varied within a list. In addition to some randomly timed filler lists, there were four critical trial types, in which the six inter-digit blank intervals were all short (0.5 s following each item) or all long (2 s following each item), or comprised three short and then three long intervals, or three long and then three short intervals. Moreover, there were two post-list response cues. According to one cue, the participant was to recall the list with the items in the presented order, but at any rate they wished. According to the other response cue, the list was to be recalled using the same timing in which it was presented. The expectation was that the need to remember the timing in the latter response condition would prevent rehearsal of either type. As a consequence, performance should be impaired on trials in which the first three response intervals are long because, on these trials, there is more time for forgetting of most of the list items. Just as predicted, there was a significant interaction between the response cue and the length of the first half of the response intervals. When participants were free to recall items at their own pace, performance was no better with a short first half ($M$=.71) than with a long first half ($M$=.74). The slight benefit of a long first half in that situation could occur because it allowed the list to be rehearsed early on in the response. In contrast, when the timing of recall had to match the timing of the list presentation, performance was better with a short first half ($M$ = .70) than with a long first half ($M$ = .67). This, then, suggests there could be decay in short-term memory.

## Overcoming contamination from long-term retrieval

If there is more than one type of memory storage then there still is the problem of which store provided the information underlying a response. There is no guarantee that, just because a

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

procedure is considered a test of short-term storage, the long-term store will not be used. For example, in a simple digit span task, a series of digits is presented and is to be repeated immediately afterward from memory. If that series turned out to be only slightly different from the participant's telephone number, the participant might be able to memorize the new number quickly and repeat it from long-term memory. The dual-store theories of memory allow this. Although Broadbent (1958) and Atkinson and Shiffrin (1968) drew their models of information processing as a series of boxes representing different memory stores, with long-term memory following short-term memory, these boxes do not imply that memory is exclusively in one box or another; they are better interpreted as the relative times of the first entry of information from a stimulus into one store and then the next. The question remains, then as to how one can determine if a response comes from short-term memory.

Waugh and Norman (1965) developed a mathematical model to accomplish this. The model operated with the assumption that long-term memory occurs for the entire list, including a plateau in the middle of the list. In contrast, by the time of recall, short-term memory is said to remain only at the end of the list. This model assumes that, for any particular serial position within a list, the likelihood of successful short-term storage (S) and long-term storage (L) are independent, so that the likelihood of recalling the item is $S+L-SL$.

A slightly different assumption is that short- and long-term stores are not independent but are used in a complementary fashion. The availability of short-term memory of an item may allow resources needed for long-term memorization to be shifted to elsewhere in the list. The data seem more consistent with that assumption. In several studies, lists to be recalled have been presented to patients with Korsakoff's amnesia and normal control participants (Baddeley and Warrington, 1970; Carlesimo et al., 1995). These studies show that, in immediate recall, performance in amnesic individuals is preserved at the last few serial positions of the list. It is as if the performance in those serial positions is based mostly or entirely on short-term storage, and that there is no decrease in that kind of storage in the amnesic patients. In delayed recall, the amnesic patients show a deficit at all serial positions, as one would expect if short-term memory for the end of the list is lost as a function of a filled delay period (Glanzer and Cunitz, 1966).

## Overcoming contamination from temporal distinctiveness

Last, it has been argued that the loss of memory over time is not necessarily the result of decay. Instead, it can be caused by temporal distinctiveness in retrieval. This kind of theory assumes that the temporal context of an item serves as a retrieval cue for that item, even in free recall. An item separated in time from all other items is relatively distinctive and easy to recall, whereas an item that is relatively close to other items is more difficult to recall because it shares their temporal cues to retrieval. Shortly after a list is presented the most recent items are the most distinct temporally (much like the distinctness of a telephone pole you are practically touching compared to poles extending further down the road). Across a retention interval, the relative distinctiveness of the most recent items decreases (much like standing far away from even the last pole in a series).

Although there are data that can be interpreted according to distinctiveness, there also are what look like dissociations between the effects of distinctiveness and a genuine short-term memory effect. One can see this, for example, in the classic procedure of Peterson and Peterson (1959) in which letter trigrams are to be recalled immediately or only after a distracting task, counting backward from a starting number by three, for a period lasting up to 18 s. Peterson and Peterson found severe memory loss for the letter trigram as the filled delay was increased. However, subsequently, sceptics argued that the memory loss occurred because the temporal distinctiveness of the current letter trigram diminished as the filled delay increased. In particular, this delay effect was said to occur because of the increase across test delays in the

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

proactive interference from previous trials. On the first few trials, the delay does not matter (Keppel and Underwood, 1962) and no detrimental effect of delay is observed if delays of 5, 10, 15, and 20 s are tested in separate trial blocks (Turvey et al., 1970; Greene, 1996).

Yet, there may be a true decay effect at shorter test intervals. Baddeley and Scott (1971) set up a trailer in a shopping mall so that they could test a large number of participants for one trial each, so as to avoid proactive interference. They found an effect of the test delay within the first 5 s but not at longer delays. Still, it seems that the concept of decay is not yet on very firm ground and warrants further study. It may be that decay actually reflects not a gradual degradation of the quality of the short-term memory trace, but a sudden collapse at a point that varies from trial to trial. With a control for temporal distinctiveness, Cowan et al. (1997a) found what could be a sudden collapse in the representation of memory for a tone with delays of between 5 and 10 s.

## Chunk capacity limits

The concept of capacity limits was raised several times in the history of cognitive psychology. Miller (1956) famously discussed the “magical number seven plus or minus two” as a constant in short-term processing, including list recall, absolute judgment, and numerical estimation experiments. However, his autobiographical essay (Miller, 1989) indicates that he was never very serious about the number seven; it was a rhetorical device that he used to tie together the otherwise unrelated strands of his research for a talk. Although it is true that memory span is approximately seven items in adults, there is no guarantee that each item is a separate entity. Perhaps the most important point of Miller’s (1956) article was that multiple items can be combined into a larger, meaningful unit. Later studies suggested that the limit in capacity is more typically only three or four units (Broadbent, 1975; Cowan, 2001). That conclusion was based on an attempt to take into account strategies that often increase the efficiency of use of a limited capacity, or that allow the maintenance of additional information separate from that limited capacity. To understand these methods of discussing capacity limits I will again mention three types of contamination. These come from chunking and the use of long-term memory, from rehearsal, and from non-capacity-limited types of storage.

### Overcoming contamination from chunking and the use of long-term memory

A participant’s response in an immediate-memory task depends on how the information to be recalled is grouped to form multi-item chunks (Miller, 1956). Because it is not usually clear what chunks have been used in recall, it is not clear how many chunks can be retained and whether the number is truly fixed. Broadbent (1975) proposed some situations in which multi-item chunk formation was not a factor, and suggested on the basis of results from such procedures that the true capacity limit is three items (each serving as a single-item chunk). For example, although memory span is often about seven items, errors are made with seven-item lists and the error-free limit is typically three items. When people must recall items from a category in long-term memory, such as the states of the United States, they do so in spurts of about three items on average. It is as if the bucket of short-term memory is filled from the well of long-term memory and must be emptied before it is refilled. Cowan (2001) noted other such situations in which multi-item chunks cannot be formed. For example, in running memory span, a long list of items is presented with an unpredictable endpoint, making grouping impossible. When the list ends, the participant is to recall a certain number of items from the end of the list. Typically, people can recall three or four items from the end of the list, although the exact number depends on task demands (Bunting et al., 2006). Individuals differ in capacity, which ranges from about two to six items in adults (and fewer in children), and the individual capacity limit is a strong correlate of cognitive aptitude.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Another way to take into account the role of multi-item chunk formation is to set up the task in a manner that allows chunks to be observed. Tulving and Patkau (1962) studied free recall of word lists with various levels of structure, ranging from random words to well-formed English sentences, with several different levels of coherence in between. A chunk was defined as a series of words reproduced by the participant in the same order in which the words had been presented. It was estimated that, in all conditions, participants recalled an average of four to six chunks. Cowan et al. (2004) tried to refine that method by testing serial recall of eight-word lists, which were composed of four pairs of words that previously had been associated with various levels of learning (0, 1, 2, or 4 prior word–word pairings). Each word used in the list was presented an equal number of times (four, except in a non-studied control condition) but what varied was how many of those presentations were as singletons and how many were as a consistent pairing. The number of paired prior exposures was held constant across the four pairs in a list. A mathematical model was used to estimate the proportion of recalled pairs that could be attributed to the learned association (i.e., to a two-word chunk) as opposed to separate recall of the two words in a pair. This model suggested that the capacity limit was about 3.5 chunks in every learning condition, but that the ratio of two-word chunks to one-word chunks increased as a function of the number of prior exposures to the pairs in the list.

## Overcoming contamination from rehearsal

The issue of rehearsal is not entirely separate from the issue of chunk formation. In the traditional concept of rehearsal (e.g., Baddeley, 1986), one imagines that the items are covertly articulated in the presented order at an even pace. However, another possibility is that rehearsal involves the use of articulatory processes in order to put the items into groups. In fact, Cowan et al. (2006a) asked participants in a digit span experiment how they carried out the task and by far the most common answer among adults was that they grouped the items; participants rarely mentioned saying the items to themselves. Yet, it is clear that suppressing rehearsal affects performance.

Presumably, the situations in which items cannot be rehearsed are for the most part the same as the situations in which items cannot be grouped. For example, Cowan et al. (2005) relied on a running memory span procedure in which the items were presented at the rapid rate of 4 per second. At that rate, it is impossible to rehearse the items as they are presented. Instead, the task is probably accomplished by retaining a passive store (sensory or phonological memory) and then transferring the last few items from that store into a more attention-related store at the time of recall. In fact, with a fast presentation rate in running span, instructions to rehearse the items is detrimental, not helpful, to performance (Hockey, 1973). Another example is memory for lists that were ignored at the time of their presentation (Cowan et al., 1999). In these cases, the capacity limit is close to the three or four items suggested by Broadbent (1975) and Cowan (2001).

It is still quite possible that there is a speech-based short-term storage mechanism that is by and large independent of the chunk-based mechanism. In terms of the popular model of Baddeley (2000), the former is the phonological loop and the latter, the episodic buffer. In terms of Cowan (1988, 1995, 1999, 2005), the former is part of activated memory, which may have a time limit due to decay, and the latter is the focus of attention, which is assumed to have a chunk capacity limit.

Chen and Cowan (2005) showed that the time limit and chunk capacity limit in short-term memory are separate. They repeated the procedure of Cowan et al. (2004) in which pairs of words sometimes were presented in a training session preceding the list recall test. They combined lists composed of pairs as in that study. Now, however, both free and serial recall tasks were used, and the length of list varied. For long lists and free recall, the chunk capacity limit governed the recall. For example, lists of six well-learned pairs were recalled as well as

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

lists of six unpaired singletons (i.e., were recalled at similar proportions of words correct). For shorter lists and serial recall strictly scored, the time limit instead governed the recall. For example, lists of four well-learned pairs were not recalled nearly as well as lists of four unpaired singletons, but only as well as lists of eight unpaired singletons. For intermediate conditions it appeared as if chunk capacity limits and time limits operate together to govern recall. Perhaps the capacity-limited mechanism holds items and the rehearsal mechanism preserves some serial order memory for those held items. The exact way in which these limits work together is not yet clear.

## Overcoming contamination from non-capacity-limited types of storage

It is difficult to demonstrate a true capacity limit that is related to attention if, as I believe, there are other types of short-term memory mechanisms that complicate the results. A general capacity should include chunks of information of all sorts: for example, information derived from both acoustic and visual stimuli, and from both verbal and nonverbal stimuli. If this is the case, there should be cross-interference between one type of memory load and another. However, the literature often has shown that there is much more interference between similar types of memoranda, such as two visual arrays of objects or two acoustically presented word lists, than there is between two dissimilar types, such as one visual array and one verbal list. Cocchini et al. (2002) suggested that there is little or no interference between dissimilar lists. If so, that would appear to provide an argument against the presence of a general, cross-domain, short-term memory store.

Morey and Cowan (2004, 2005) questioned this conclusion. They presented a visual array of colored spots to be compared to a second array that matched the first or differed from it in one spot's color. Before the first array or just after it, participants sometimes heard a list of digits that was then to be recited between the two arrays. In a low-load condition, the list was their own seven-digit telephone number whereas, in a high-load condition, it was a random seven-digit number. Only the latter condition interfered with array-comparison performance, and then only if the list was to be recited aloud between the arrays. This suggests that retrieving seven random digits in a way that also engages rehearsal processes relies upon some type of short-term memory mechanism that also is needed for the visual arrays. That shared mechanism may be the focus of attention, with its capacity limit. Apparently, though, if the list was maintained silently rather than being recited aloud, this silent maintenance occurred without much use of the common, attention-based storage mechanism, so visual array performance was not much affected.

The types of short-term memory whose contribution to recall may obscure the capacity limit can include any types of activated memory that fall outside of the focus of attention. In the modeling framework depicted in Fig. 1, this can include sensory memory features as well as semantic features. Sperling (1960) famously illustrated the difference between unlimited sensory memory and capacity-limited categorical memory. If an array of characters was followed by a partial report cue shortly after the array, most of the characters in the cued row could be recalled. If the cue was delayed about 1 s, most of the sensory information had decayed and performance was limited to about four characters, regardless of the size of the array. Based on this study, the four-character limit could be seen as either a limit in the capacity of short-term memory or a limit in the rate with which information could be transferred from sensory memory into a categorical form before it decayed. However, Darwin et al. (1972) carried out an analogous auditory experiment and found a limit of about four items even though the observed decay period for sensory memory was about 4s. Given the striking differences between Sperling and Darwin et al. in the time period available for the transfer of information to a categorical form, the common four-item limit is best viewed as a capacity limit rather than a rate limit.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



Saults and Cowan (2007) tested this conceptual framework in a series of experiments in which arrays were presented in two modalities at once or, in another procedure, one after the other. A visual array of colored spots was supplemented by an array of spoken digits occurring in four separate loudspeakers, each one consistently assigned to a different voice to ease perception. On some trials, participants knew that they were responsible for both modalities at once whereas, in other trials, participants knew that they were responsible for only the visual or only the acoustic stimuli. They received a probe array that was the same as the previous array (or the same as one modality in that previous array) or differed from the previous array in the identity of one stimulus. The task was to determine if there was a change. The use of cross-modality, capacity-limited storage predicts a particular pattern of results. It predicts that performance on either modality should be diminished in the dual-modality condition compared to the unimodal conditions, due to strain on the cross-modality store. That is how the results turned out. Moreover, if the cross-modality, capacity-limited store were the only type of storage used, then the sum of visual and auditory capacities in the dual-modality condition should be no greater than the larger of the two unimodal capacities (which happened to be the visual capacity). The reason is that the limited-capacity store would hold the same number of units no matter whether they were all from one modality or were from two modalities combined. That prediction was confirmed, but only if there was a post-perceptual mask in both modalities at once following the array to be remembered. The post-perceptual mask included a multicolored spot at each visual object location and a sound composed of all possible digits overlaid, from each loudspeaker. It was presented long enough after the arrays to be recalled that their perception would have been complete (e.g., 1 s afterward; cf. Vogel et al., 2006). Presumably, the mask was capable of overwriting various types of sensory-specific features in activated memory, leaving behind only the more generic, categorical information present in the focus of attention, which presumably is protected from masking interference by the attention process. The limit of the focus of attention was again shown to be between three and four items, for either unimodal visual or bimodal stimuli.

Even without using masking stimuli, it may be possible to find a phase of the short-term memory process that is general across domains. Cowan and Morey (2007) presented two stimulus sets to be recalled (or, in control conditions, only one set). The two stimulus sets could include two spoken lists of digits, two spatial arrays of colored spots, or one of each, in either order. Following this presentation, a cue indicated that the participant would be responsible for only the first array, only the second array, or both arrays. Three seconds followed before a probe. The effect of memory load could be compared in two ways. Performance on those trials in which two sets of stimuli were presented and both were cued for retention could be compared either to trials in which only one set was presented, or it could be compared to trials in which both sets were presented but the cue later indicated that only one set had to be retained. The part of working memory preceding the cue showed modality-specific dual-task effects: encoding a stimulus set of one type was hurt more by also encoding another set if both sets were in the same modality. However, the retention of information following the cue showed dual-task effects that were not modality-specific. When two sets had been presented, retaining both of them was detrimental compared to retaining only one set (as specified by the post-stimulus retention cue to retain one versus both sets), and this dual-task effect was similar in magnitude no matter whether the sets were in the same or different modalities. After the initial encoding, working memory storage across several seconds thus may occur abstractly, in the focus of attention.

## Other evidence for a separate short-term storage

Last, there is other evidence that does not directly support either temporal decay or a capacity limit specifically, but implies that one or the other of these limits exist. Bjork and Whitten (1974) and Tzeng (1973) made temporal distinctiveness arguments on the basis of what is

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

called continual distractor list recall, in which a recency effect persists even when the list is followed by a distracter-filled delay before recall. The filled delay should have destroyed short-term memory but the recency effect occurs anyway, provided that the items in the list also are separated by distracter-filled delays to increase their distinctiveness from one another. In favor of short-term storage, though, other studies have shown dissociations between what is found in ordinary immediate recall versus continual distractor recall (e.g., word length effects reversed in continual distractor recall: Cowan et al., 1997b; proactive interference at the most recent list positions in continual distractor recall only: Craik & Birtwistle, 1971; Davelaar et al., 2005).

There is also additional neuroimaging evidence for short-term storage. Talmi et al. (2005) found that recognition of earlier portions of a list, but not the last few items, activated areas within the hippocampal system that is generally associated with long-term memory retrieval. This is consistent with the finding, mentioned earlier, that memory for the last few list items is spared in Korsakoff’s amnesia (Baddeley and Warrington, 1970; Carlesimo et al., 1995). In these studies, the part of the recency effect based on short-term memory could reflect a short amount of time between presentation and recall of the last few items, or it could reflect the absence of interference between presentation and recall of the last few items. Thus, we can say that short-term memory exists, but often without great clarity as to whether the limit is a time limit or a chunk capacity limit.

## The short-term memory/working memory distinction

The distinction between short-term memory and working memory is clouded in a bit of confusion but that is largely the result of different investigators using different definitions. Miller et al. (1960) used the term “working memory” to refer to temporary memory from a functional standpoint, so from their point of view there is no clear distinction between short-term and working memory. Baddeley and Hitch (1974) were fairly consistent with this definition but overlaid some descriptions on the terms that distinguished them. They thought of short-term memory as the unitary holding place as described by, for example, Atkinson and Shiffrin (1968). When they realized that the evidence actually was consistent with a multi-component system that could not be reduced to a unitary short-term store, they used the term working memory to describe that entire system. Cowan (1988) maintained a multi-component view, like Baddeley and Hitch, but without a commitment to precisely their components; instead, the basic subdivisions of working memory were said to be the short-term storage components (activated memory along with the focus of attention within it, shown in Fig. 1) and central executive processes that manipulate stored information. By Cowan’s account, Baddeley’s (1986) phonological loop and visuospatial sketchpad would be viewed as just two of many aspects of activated memory, which are susceptible to interference to a degree that depends upon the similarity between features of the activated and interfering information sources. Baddeley’s (2000) episodic buffer is possibly the same as the information saved in Cowan’s focus of attention, or at least is a closely similar concept.

There has been some shift in the definition or description of working memory along with a shift in the explanation of why the newer working memory tasks correlate with intelligence and aptitude measures so much more highly than do simple, traditional, short-term memory tasks such as serial recall. Daneman and Carpenter (1980) had assumed that what is critical is to use working memory tasks that include both storage and processing components, so as to engage all of the parts of working memory as described, for example, by Baddeley and Hitch (1974). Instead, Engle et al. (1999) and Kane et al. (2001) proposed that what is critical is whether the working memory task is challenging in terms of the control of attention. For example, Kane et al. found that working memory span storage-and-processing tasks correlates well with the ability to inhibit the natural tendency to look toward a suddenly appearing

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

stimulus and instead to look the other way, the antisaccade task. Similarly, Conway et al. (2001) found that individuals scoring high on storage-and-processing tests of working memory notice their names in a channel to be ignored in dichotic listening much *less* often than low-span individuals; the high-span individuals apparently are better able to make their primary task performance less vulnerable to distraction, but this comes at the expense of being a bit oblivious to irrelevant aspects of their surroundings. In response to such research, Engle and colleagues sometimes used the term working memory to refer only to the processes related to controlling attention. By doing so, their definition of working memory seems at odds with previous definitions but that new definition allows the simple statement that working memory correlates highly with aptitudes, whereas short-term memory (redefined to include only the non-attention-related aspects of memory storage) does not correlate so highly with aptitudes.

Cowan et al. (2006b), while adhering to the more traditional definition of working memory, made an assertion about working memory similar to that of Engle and colleagues, but a bit more complex. They proposed, on the basis of some developmental and correlational evidence, that multiple functions of attention are relevant to individual differences in aptitudes. The control of attention is relevant, but there is an independent contribution from the number of items that can be held in attention, or its scope. According to this view, what may be necessary for a working memory procedure to correlate well with cognitive aptitudes is that the task must prevent covert verbal rehearsal so that the participant must rely on more attention-demanding processing and/or storage to carry out the task. Cowan et al. (2005) found that the task can be much simpler than the storage-and-processing procedures. For example, in a version of the running memory span test, digits are presented very quickly and the series stops at an unpredictable point, after which the participant is to recall as many items as possible from the end of the list. Rehearsal is impossible and, when the list ends, information presumably must be retrieved from activated sensory or phonological features into the focus of attention. This type of task correlated with aptitudes, as did several other measures of the scope of attention (Cowan et al., 2005, 2006b). In children too young to use covert verbal rehearsal (unlike older children and adults), even a simple digit span task served as an excellent correlate with aptitudes.

Other research verifies this idea that a working memory test will correlate well with cognitive aptitudes to the extent that it requires that attention be used for storage and/or processing. Gavens and Barrouillet (2004) carried out a developmental study in which they controlled the difficulty and duration of a processing task that came between items to be recalled. There still was a developmental difference in span, which they attributed to the development of a basic capacity, which could reflect a developmental increase in the scope of attention (cf. Cowan et al., 2005). Lépine et al. (2005) showed that what was important for a storage-and-processing type of span task to correlate well with aptitudes is for the processing component of the task (in this case, reading letters aloud) to occur quickly enough to prevent various types of rehearsal to sneak in between (see also Conlin et al., 2005).

Several papers have pitted storage and processing (perhaps the scope versus control of attention?) against one another to see which is more important in accounting for individual differences. Vogel et al. (2005) used a visual array task modified for use with a component of event-related potentials that indicates storage in visual working memory, termed contralateral delay activity (CDA). This activity was found to depend not only on the number of relevant objects in the display (e.g., red bars at varying angles to be remembered), but sometimes also on the number of irrelevant objects to be ignored (e.g., blue bars). For high-span individuals, the CDA for two relevant objects was found to be similar whether or not there also were two irrelevant objects in the display. However, for low-span individuals, the CDA for two relevant objects combined with two irrelevant objects was similar to the CDA for displays with four relevant objects alone, as if the irrelevant objects could not be excluded from working memory.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

One limitation of the study is that the separation of participants into high versus low span was based on the CDA also, and the task used to measure the CDA inevitably required selective attention (to one half of the display) on every trial, whether or not it included objects of an irrelevant color.

Gold et al. (2006) investigated similar issues in a behavioral design, and testing the difference between schizophrenic patients and normal control participants. Each trial started with a cue to attend to one part of the display at the expense of another (e.g., bars of one relevant color but not another, irrelevant color). The probe display was a set that was cued for relevance on most trials (in some experiments, 75%) whereas, occasionally, the probe display was a set that was not cued. This allowed a separate measure of the control of attention (the advantage for cued items over uncued items) and the storage capacity of working memory (the mean number of items recalled from each array, adding across cued and uncued sets). Unlike the initial expectations, the clear result was that the difference between groups was in the capacity, not in the control of attention. It would be interesting to know whether the same type of result could be obtained for high versus low span normal individuals, or whether that comparison instead would show a control-of-attention difference between these groups as Vogel et al. (2005) must predict. Friedman et al. (2006) found that not all central executive functions correlated with aptitudes; updating working memory did, but inhibition and shifting of attention did not. On the other hand, recall that Cowan et al. (2006b) did find was that a control-of-attention task was related to aptitudes.

In sum, the question of whether short-term memory and working memory are different may be a matter of semantics. There are clearly differences between simple serial recall tasks that do not correlate very well with aptitude tests in adults, and other tasks requiring memory and processing, or memory without the possibility of rehearsal, that correlate much better with aptitudes. Whether to use the term working memory for the latter set of tasks, or whether to reserve that term for the entire system of short-term memory preservation and manipulation, is a matter of taste. The more important, substantive question may be why some tasks correlate with aptitude much better than others.

## Conclusion

The distinction between long-term and short-term memory depends on whether it can be demonstrated that there are properties specific to short-term memory; the main candidates include temporal decay and a chunk capacity limit. The question of decay is still pretty much open to debate, whereas there is growing support for a chunk capacity limit. These limits were discussed in a framework shown in Fig. 1.

The distinction between short-term memory and working memory is one that depends on the definition that one accepts. Nevertheless, the substantive question is why some tests of memory over the short term serve as some of the best correlates of cognitive aptitudes, whereas others do not. The answer seems to point to the importance of an attentional system used both for processing and for storage. The efficiency of this system and its use in working memory seem to differ substantially across individuals (e.g., Conway et al., 2002; Kane et al., 2004; Cowan et al., 2005, 2006b), as well as improving with development in childhood (Cowan et al., 2005, 2006b) and declining in old age (Naveh-Benjamin et al., 2007; Stoltzfus et al., 1996; Cowan et al., 2006c).

## Acknowledgment

This work was completed with the assistance of NIH Grant R01 HD-21338.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

## References


Atkinson, RC.; Shiffrin, RM. Human memory: a proposed system and its control processes. In: Spence, KW.; Spence, JT., editors. The Psychology of Learning and Motivation: Advances in Research and Theory. Vol. Vol. 2. New York: Academic Press; 1968. p. 89-195.

Baddeley A. The episodic buffer: a new component of working memory? Trends Cogn. Sci 2000;4:417–423. [PubMed: 11058819]

Baddeley, AD. Oxford Psychology Series No. 11. Oxford: Clarendon Press; 1986. Working Memory.

Baddeley, AD.; Hitch, G. Working memory. In: Bower, GH., editor. The Psychology of Learning and Motivation. Vol. Vol. 8. New York: Academic Press; 1974. p. 47-89.

Baddeley AD, Scott D. Short-term forgetting in the absence of proactive inhibition. Q. J. Exp. Psychol 1971;23:275–283.

Baddeley AD, Thomson N, Buchanan M. Word length and the structure of short-term memory. J. Verbal Learn. Verbal Behav 1975;14:575–589.

Baddeley AD, Warrington EK. Amnesia and the distinction between long- and short-term memory. J. Verbal Learn. Verbal Behav 1970;9:176–189.

Barrouillet P, Bernardin S, Camos V. Time constraints and resource sharing in adults' working memory spans. J. Exp. Psychol.: Gen 2004;133:83–100. [PubMed: 14979753]

Barrouillet P, Bernardin S, Portrat S, Vergauwe E, Camos V. Time and cognitive load in working memory. J. Exp. Psychol. Learn. Mem. Cogn 2007;33:570–585. [PubMed: 17470006]

Bjork RA, Whitten WB. Recency-sensitive retrieval processes in long-term free recall. Cogn. Psychol 1974;6:173–189.

Broadbent, DE. Perception and Communication. New York: Pergamon Press; 1958.

Broadbent, DE. The magic number seven after fifteen years. In: Kennedy, A.; Wilkes, A., editors. Studies in Long-Term Memory. Oxford, England: Wiley; 1975. p. 3-18.

Brown GDA, Preece T, Hulme C. Oscillator-based memory for serial order. Psychol. Rev 2000;107:127–181. [PubMed: 10687405]

Bunting MF, Cowan N, Saults JS. How does running memory span work? Q. J. Exp. Psychol 2006;59:1691–1700.

Carlesimo GA, Sabbadini M, Fadda L, Caltagirone C. Different components in word-list forgetting of pure amnesics, degenerative demented and healthy subjects. Cortex 1995;31:735–745. [PubMed: 8750030]

Chen Z, Cowan N. Chunk limits and length limits in immediate recall: a reconciliation. J. Exp. Psychol. Learn. Mem. Cogn 2005;31:1235–1249. [PubMed: 16393043]

Cocchini G, Logie RH, Della Sala S, MacPherson SE, Baddeley AD. Concurrent performance of two memory tasks: evidence for domain-specific working memory systems. Mem. Cogn 2002;30:1086–1095.

Conlin JA, Gathercole SE, Adams JW. Children's working memory: investigating performance limitations in complex span tasks. J. Exp. Child Psychol 2005;90:303–317. [PubMed: 15777923]

Conway ARA, Cowan N, Bunting MF. The cocktail party phenomenon revisited: the importance of working memory capacity. Psychon. Bull. Rev 2001;8:331–335. [PubMed: 11495122]

Conway ARA, Cowan N, Bunting MF, Therriault DJ, Minkoff SRB. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. Intelligence 2002;30:163–183.

Conway ARA, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: a methodological review and user's guide. Psychon. Bull. Rev 2005;12:769–786. [PubMed: 16523997]

Cowan N. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. Psychol. Bull 1988;104:163–191. [PubMed: 3054993]

Cowan N. Verbal memory span and the timing of spoken recall. J. Mem. Lang 1992;31:668–684.

Cowan, N. Oxford Psychology Series No. 26. New York: Oxford University Press; 1995. Attention and memory: an integrated framework.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Cowan, N. An embedded-processes model of working memory. In: Miyake, A.; Shah, P., editors. Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. Cambridge, UK: Cambridge University Press; 1999. p. 62-101.

Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav. Brain Sci 2001;24:87–185. [PubMed: 11515286]

Cowan, N. Working memory capacity. Hove, East Sussex, UK: Psychology Press; 2005.

Cowan, N.; Aubuchon, AM. Psychon. Bull. Rev. Short-term memory loss over time without retroactive stimulus interference. (in press)

Cowan N, Chen Z, Rouder JN. Constant capacity in an immediate serial-recall task: a logical sequel to Miller (1956). Psychol. Sci 2004;15:634–640. [PubMed: 15327636]

Cowan N, Elliott EM, Saults JS, Morey CC, Mattox S, Hismjatullina A, Conway ARA. On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. Cogn. Psychol 2005;51:42–100. [PubMed: 16039935]

Cowan N, Elliott EM, Saults JS, Nugent LD, Bomb P, Hismjatullina A. Rethinking speed theories of cognitive development: increasing the rate of recall without affecting accuracy. Psychol. Sci 2006a; 17:67–73. [PubMed: 16371146]

Cowan N, Fristoe NM, Elliott EM, Brunner RP, Saults JS. Scope of attention, control of attention, and intelligence in children and adults. Mem. Cogn 2006b;34:1754–1768.

Cowan N, Morey CC. How can dual-task working memory retention limits be investigated? Psychol. Sci 2007;18:686–688. [PubMed: 17680938]

Cowan N, Naveh-Benjamin M, Kilb A, Saults JS. Life-span development of visual working memory: when is feature binding difficult? Dev. Psychol 2006c;42:1089–1102. [PubMed: 17087544]

Cowan N, Nugent LD, Elliott EM, Ponomarev I, Saults JS. The role of attention in the development of short-term memory: age differences in the verbal span of apprehension. Child Dev 1999;70:1082–1097. [PubMed: 10546336]

Cowan N, Saults JS, Nugent LD. The role of absolute and relative amounts of time in forgetting within immediate memory: the case of tone pitch comparisons. Psychon. Bull. Rev 1997a;4:393–397.

Cowan N, Wood NL, Nugent LD, Treisman M. There are two word length effects in verbal short-term memory: opposed effects of duration and complexity. Psychol. Sci 1997b;8:290–295.

Craik FIM, Birtwistle J. Proactive inhibition in free recall. J. Exp. Psychol 1971;91:120–123.

Crowder RG. The demise of short-term memory. Acta Psychol 1982;50:291–323.

Crowder RG. Short-term memory: where do we stand? Mem. Cogn 1993;21:142–145.

Daneman M, Carpenter PA. Individual differences in working memory and reading. J Verbal Learn. Verbal Behav 1980;19:450–466.

Daneman M, Merikle PM. Working memory and language comprehension: a meta-analysis. Psychon. Bull. Rev 1996;3:422–433.

Darwin CJ, Turvey MT, Crowder RG. An auditory analogue of the Sperling partial report procedure: evidence for brief auditory storage. Cogn. Psychol 1972;3:255–267.

Davelaar EJ, Goshen-Gottstein Y, Ashkenazi A, Haarman HJ, Usher M. The demise of short-term memory revisited: empirical and computational investigations of recency effects. Psychol. Rev 2005;112:3–42. [PubMed: 15631586]

Ebbinghaus, H. Translated by H.A. Ruger and C.E. Bussenius. New York: Teachers College, Columbia University; 18851913. Memory: a contribution to experimental psychology. (Originally in German, Ueber das gedächtnis: Untersuchen zur experimentellen psychologie)

Engle RW. Working memory capacity as executive attention. Curr. Dir. Psychol. Sci 2002;11:19–23.

Engle RW, Tuholski SW, Laughlin JE, Conway ARA. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. J. Exp. Psychol. Gen 1999;128:309–331. [PubMed: 10513398]

Ericsson KA, Kintsch W. Long-term working memory. Psychol. Rev 1995;102:211–245. [PubMed: 7740089]

Friedman NP, Miyake A, Corley RP, Young SE, DeFries JC, Hewitt JK. Not all executive functions are related to intelligence. Psychol. Sci 2006;17:172–179. [PubMed: 16466426]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Gavens N, Barrouillet P. Delays of retention, processing efficiency, and attentional resources in working memory span development. J. Mem. Lang 2004;51:644–657.

Glanzer M, Cunitz AR. Two storage mechanisms in free recall. J. Verbal Learn. Verbal Behav 1966;5:351–360.

Glenberg AM, Swanson NC. A temporal distinctiveness theory of recency and modality effects. J. Exp. Psychol. Learn. Mem. Cogn 1986;12:3–15. [PubMed: 2949048]

Gold JM, Fuller RL, Robinson BM, McMahon RP, Braun EL, Luck SJ. Intact attentional control of working memory encoding in schizophrenia. J. Abnorm. Psychol 2006;115:658–673. [PubMed: 17100524]

Greene RL. The influence of experimental design: the example of the Brown-Peterson paradigm. Can. J. Exp. Psychol 1996;50:240–242.

Guttentag RE. The mental effort requirement of cumulative rehearsal: a developmental study. J. Exp. Child Psychol 1984;37:92–106.

Hebb, DO. Organization of Behavior. New York: Wiley; 1949.

Hockey R. Rate of presentation in running memory and direct manipulation of input-processing strategies. Q. J. Exp. Psychol. A 1973;25:104–111.

James, W. The Principles of Psychology. New York: Henry Holt; 1890.

Kane MJ, Bleckley MK, Conway ARA, Engle RW. A controlled-attention view of working-memory capacity. J. Exp. Psychol. Gen 2001;130:169–183. [PubMed: 11409097]

Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, Engle RE. The generality of working-memory capacity: a latent-variable approach to verbal and visuo-spatial memory span and reasoning. J. Exp. Psychol. Gen 2004;133:189–217. [PubMed: 15149250]

Keppel G, Underwood BJ. Proactive inhibition in short-term retention of single items. J. Verbal Learn. Verbal Behav 1962;1:153–161.

Kyllonen PC, Christal RE. Reasoning ability is (little more than) working-memory capacity? Intelligence 1990;14:389–433.

Lépine R, Barrouillet P, Camos V. What makes working memory spans so predictive of high level cognition? Psychon. Bull. Rev 2005;12:165–170. [PubMed: 15945209]

Lewandowsky S, Duncan M, Brown GDA. Time does not cause forgetting in short-term serial recall. Psychon. Bull. Rev 2004;11:771–790. [PubMed: 15732687]

McGeoch JA. Forgetting and the law of disuse. Psychol. Rev 1932;39:352–370.

Melton AW. Implications of short-term memory for a general theory of memory. J. Verbal Learn. Verbal Behav 1963;2:1–21.

Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev 1956;63:81–97. [PubMed: 13310704]

Miller, GA. George A. Miller. In: Lindzey, G., editor. A History of Psychology in Autobiography. Vol. Vol. VIII. Stanford, CA: Stanford University Press; 1989. p. 391-418.

Miller, GA.; Galanter, E.; Pribram, KH. Plans and the structure of behavior. New York: Holt, Rinehart and Winston, Inc; 1960.

Morey CC, Cowan N. When visual and verbal memories compete: evidence of cross-domain limits in working memory. Psychon. Bull. Rev 2004;11:296–301. [PubMed: 15260196]

Morey CC, Cowan N. When do visual and verbal memories conflict? The importance of working-memory load and retrieval. J. Exp. Psychol. Learn. Mem. Cogn 2005;31:703–713. [PubMed: 16060774]

Nairne JS. Remembering over the short-term: the case against the standard model. Annu. Rev. Psychol 2002;53:53–81. [PubMed: 11752479]

Naveh-Benjamin M, Cowan N, Kilb A, Chen Z. Age-related differences in immediate serial recall: dissociating chunk formation and capacity. Mem. Cognit 2007;35:724–737.

Neath, I.; Surprenant, A. Human memory. Vol. 2nd ed.. Belmont, CA: Wadsworth; 2003.

Nipher FE. On the distribution of errors in numbers written from memory. Trans. Acad. Sci. St. Louis 1878;3:ccx–ccxi.

Peterson LR, Peterson MJ. Short-term retention of individual verbal items. J. Exp. Psychol 1959;58:193–198. [PubMed: 14432252]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript


Saults JS, Cowan N. A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. J. Exp. Psychol 2007;136:663–684.

Sperling G. The information available in brief visual presentations. Psychol. Monogr 1960;74(Whole No. 498.)

Stoltzfus, ER.; Hasher, L.; Zacks, RT. Working memory and retrieval: an inhibition-resource approach. In: Richardson, JTE.; Engle, RW.; Hasher, L.; Logie, RH.; Stoltzfus, ER.; Zacks, RT., editors. Working Memory and Human Cognition. New York: Oxford University Press; 1996. p. 66-88.

Talmi D, Grady CL, Goshen-Gottstein Y, Moscovitch M. Neuroimaging the serial position curve: a test of single-store versus dual-store models. Psychol. Sci 2005;16:716–723. [PubMed: 16137258]

Tulving E, Patkau JE. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. Can. J. Psychol 1962;16:83–95. [PubMed: 13923055]

Turvey MT, Brick P, Osborn J. Proactive interference in short-term memory as a function of prior-item retention interval. Q. J. Exp. Psychol 1970;22:142–147.

Tzeng OJL. Positive recency effect in a delayed free recall. J. Verbal Learn. Verbal Behav 1973;12:436–439.

Unsworth N, Engle RW. The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. Psychol. Rev 2007;114:104–132. [PubMed: 17227183]

Vogel EK, McCollough AW, Machizawa MG. Neural measures reveal individual differences in controlling access to working memory. Nature 2005;438:500–503. [PubMed: 16306992]

Vogel EK, Woodman GF, Luck SJ. The time course of consolidation in visual working memory. J. Exp. Psychol. Hum. Percept. Perform 2006;32:1436–1451. [PubMed: 17154783]

Waugh NC, Norman DA. Primary memory. Psychol. Rev 1965;72:89–104. [PubMed: 14282677]

Wickelgren WA. Single-trace fragility theory of memory dynamics. Mem. Cogn 1974;2:775–780.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

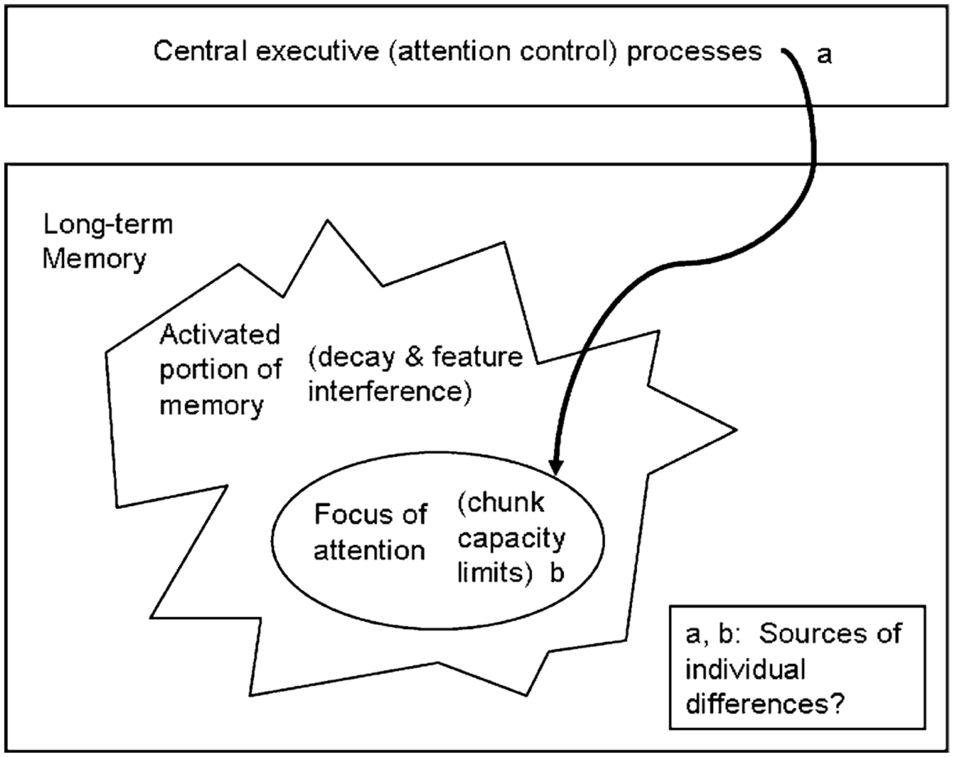


**Fig. 1.**
A depiction of the theoretical modeling framework. Modified from Cowan (1988) and refined in further work by Cowan (1995, 1999, 2005).