Working Memory

Barbara A. Dosher, University of California, Irvine, California, USA

CONTENTS
- The role of working memory in information processing
- Working memory framework
- Forms of working memory
- Models of serial memory
- Biological substrates of working memory
- Working memory, short-term memory, attention, and executive function
- Working memory and general cognitive performance

Working memory is the capacity to manipulate and maintain information over short periods (2–15 seconds) in order to support simple memory tasks such as remembering a telephone number, or more general cognitive tasks such as problem solving, simple reasoning, or reading. Working memory consists of several distinguishable memory capacities, together with executive functions that manage information retrieval, reactivation, and transformation.

THE ROLE OF WORKING MEMORY IN INFORMATION PROCESSING

Human memory—that is, the retention of information over short or long delays—is critically important for cognitive function. Working memory is the capacity to manipulate and maintain information over short temporal delays. Working memory supports cognitive activities such as solving mathematical problems, evaluating spatial layouts, or comprehending sentences, and it may be critical in language learning. It is one of several closely related short-duration memory systems. Working memory is of longer duration than the relatively unprocessed and very-short-duration (<2 seconds) visual sensory memory ( iconic memory) and auditory sensory memory (echoic memory). It maintains information over a period of several seconds or longer through proactive control and rehearsal mechanisms. Thus working memory goes beyond the short-term or immediate memory function that briefly maintains information to include attention and information management—the capacity for manipulation and transformation of information that is required during cognitive tasks.

WORKING MEMORY FRAMEWORK

According to one classic view, working memory is a system of three interacting modules (Baddeley, 1986). These include an executive control module, a visual–spatial sketchpad, and an articulatory loop (Figure 1). Evidence for the separable visual and verbal modules arises from the differential impact of visual distraction on memory for visual information, and the differential impact of verbal distraction on memory for verbal information. The executive control module manages task goals, manipulation of information, and complex rehearsal. The visual–spatial sketchpad maintains visual information (visual content) and spatial information (spatial layout). The articulatory loop maintains temporally ordered verbal information. The sketchpad and the loop are subservient to the executive control module.

The modular working memory framework was originally developed partly to account for capacity limitations in verbal tasks that require correctly ordered recall. In verbal recall tasks, a written or spoken list of verbal items is presented in temporal sequence, and the list is then recalled in order. Several phenomena in serial ordered recall were critical in the development of the modular working memory framework. First, lists of phonologically more dissimilar or distinct items are recalled more accurately than lists of phonologically similar items. This phonological similarity effect implies that a phonological or articulatory representation is dominant in working memory. Secondly, the number of items that can be recalled depends upon word length. Longer lists (up to 7–9 items) can be correctly recalled if those items are 'short', with a smaller number of syllables or a shorter time required to pronounce them, but recall is reduced for lists of items that are 'long' (a word-length effect). The sensitivity of working memory performance to word length suggested the existence of a cyclic rehearsal loop in which items were refreshed or rehearsed in turn in order to maintain memory. This emphasizes the importance of control processes. Thirdly, the repeated speaking of a distracting word or phrase during the memory task dramatically reduces the length of list that can be
Figure 1. The modular working memory model (Baddeley, 1986) consists of three modules. The executive control module manages rehearsal and other executive functions to transform and maintain information in working memory. The visual–spatial sketchpad maintains visual (content) and spatial (layout) information for forms, patterns, and faces. The articulatory loop maintains item and order information through verbal, articulatory codes for digits, letters, words, and sentences.

recalled. This is known as the articulatory suppression effect. Articulatory suppression also eliminates phonological similarity effects for visual presentation.

These phenomena and their complex interplay suggested the operation of an articulatory/phono-logical rehearsal mechanism, verbal recoding of visual inputs, and the importance of control or rehearsal activities—all critical concepts in the modular working memory framework. For a review of these and related phenomena see Neath (1998).

Other alternative frameworks and computational models based on quite different principles may also account for a wide range of these data. One alternative view inspired by human cognitive neuroscience and neural network representations is that short-term memory is the currently active subset of long-term memory, and that working memory is the system of short-term memory plus the strategy and attention functions that reactivate fading short-term memory traces (Cowan, 1995). Other formal models have been developed for several specific tasks (Page and Norris, 1998; Henson, 2001), or to specify more precisely the representation of the abstract and modality-specific features of items in working memory (Neath, 1998). As the brain areas relevant to working memory function are identified, and the temporal properties of the responses in these regions are more fully understood, the description of the modular working memory framework may be transformed into a more detailed specification of the computational implementation and the brain circuitry, consisting of processing structures and more complex computational network representations. Nonetheless, the modular working memory architecture (Baddeley, 1986) provides a structural-level description of memory function, and has provided a framework for an extensive body of behavioral and physiological studies of working memory in a range of memory tasks. The important distinctions between classes of working memory tasks will be considered in the following sections.

FORMS OF WORKING MEMORY

Working memory operates somewhat differently in tasks with differing complexity or cognitive processing demands. One task that is used extensively in animal models and in physiological analyses requires a single item to be maintained through a delay period. More demanding working memory tasks involve the full recall of a list of items during complex distracting activities (Daneman and Carpenter, 1980). Some tasks focus on visual coding, while others focus on verbal coding. Each form of working memory task places different demands on information retrieval and executive function.

Working memory tasks are of several major types (Figure 2). Item memory tasks require the evaluation of an individual item or individual position, although the memory set may consist of multiple items or positions. Relational memory tasks require the evaluation of some relational property, such as relative temporal order or the spatial location of several items. Finally, complex memory tasks require extended reproduction of a memory set with a sequence of temporally or spatially ordered responses. Item, relational, and complex tasks vary further in difficulty depending on the level of the memory load or of distraction.

Item Working Memory Tasks

Item working memory tasks test memory for a single item, possibly from among a set of items, following a brief delay. In the simplest item task (Figure 2(a)), recognition memory for a single item is tested after a retention delay, usually of between 1 and 60 seconds (Figure 3(a)). Recognition of the item declines with increasing delay. This task has been important for studying short-term retention in animals (the delayed match to sample task). It has also been important in studies of delay period activity of cortical neurons, and sustained neural activity during the retention interval, that has been associated with working memory (Chafee and Goldman-Rakic, 1998). For humans, delayed
**Figure 2.** Item, relational, and complex working memory tasks are successively more demanding of working memory. (a and b) Examples of item working memory tasks. (a) The delayed recognition task (a simple item working memory task) measures forgetting over short time delays. In this example, the stimuli require visual rather than verbal coding. (b) Item recognition with memory load task requires memory for an individual item from a memory set, or recognition after filled delays. In this example, the stimuli utilize verbal coding. (c and d) Examples of relational (order) working memory tasks. (c) In the temporal order task, the most recent of two items from a working memory set is selected. (d) In the n-back task, an item repetition is detected at a specified delay, 1-back (immediate repetition), 2-back (repetition with one intervening item). (e and f) Examples of complex working memory tasks. (e) In memory span tasks, items are recalled in the correct order. (f) In running span tasks, identified items from a series of mental tasks (e.g., simple mathematical problems) are recalled in the correct order. Reproduced with permission of B. Dosher.
Figure 3. Item tasks are characterized by simple decay functions and load-independent retrieval. Schematic results of a delayed item recognition task are shown. (a) Mean reaction time (RT) to recognize a single item in working memory increases with the memory load or list length. These aggregate results reflect faster response times for more recent list items. (b) Reaction time (and accuracy) of item recognition depends upon recency. Response times are shortest for list-final (most recent) positions, and they increase as items become less recent. (c) Recognition accuracy (bias-free strength $d'$) is measured as a function of retrieval time by interrupting recognition after various amounts of time for working memory loads (list lengths) of from 4 to 6 items (data from McElree and Dosher, 1989). Limiting accuracy is higher for more recent items. However, the time course of item recognition reflects parallel processing — it does not depend on working memory load or item recency. Reproduced with permission of B. Dosher.
recognition of a single study item may be too simple to tax working memory, especially for simple verbal items, so task complexity is increased by increasing the memory load, or the number of items to be remembered. Alternatively, the complexity or demand of the memory load might be increased.

One of the most extensively studied working memory tasks tests recognition of one item from a multi-item memory load presented over time (Figure 2(b)). Item recognition takes place immediately following a study list when error rates are relatively low, and response time is used to index memory. The memory load may consist of new items, in which case the judgment may reflect familiarity, or it may consist of highly repeated items from a small set, in which case the judgment must reflect not familiarity with the items themselves, but specific list membership. The average time taken to recognize an item as a member of the memory load increases approximately linearly with the load size (Figure 3(a)), which has often been claimed to reflect a serial process in retrieval from working memory (Stemberg, 1975). Recent evidence instead suggests that in simple item recognition tests, familiarity or activation is processed fairly directly, with accuracy and activation times reflecting item recency (the delay interval). Memory lists of different sizes reflect different mixtures of recency effects on both recognition time and accuracy (Figure 3(b)). In fact, under many conditions, retrieval of single items from working memory is achieved with the same time course regardless of the size of the working memory load or the study position of the item tested for recognition (Figure 3(c)). Information starts to become available at the same time and increases at the same rate independent of working memory load or temporal position (Dosher and McElree, 1992). The slower response times for larger working memory loads reflect lower memory strength for less recent items.

Simple, relatively rapid mnemonic evaluation of single items, independent of memory load, is a characteristic of item working memory. Item working memory tasks may exhibit parallel retrieval independent of the memory load. In contrast, the patterns of retrieval in relational working memory tasks exhibit substantial load effects and slower and load-dependent retrieval.

Relational Working Memory Tasks

Relational working memory tasks involve relational judgments, such as judgments of temporal or spatial order, over two or more items. These judgments place greater demands upon retrieval and rehearsal than do simple item tasks. For example, the relative order task (Figure 2(c)) requires the selection of the most recent of two test items. In comparison with the relatively rapid and load-independent retrieval exhibited for item tasks, processing of the relative temporal order (recency) is considerably slower than item retrieval, and is strongly dependent on the temporal positions of the items. The time course of judgments is rapid for memory probes including the most recent item, perhaps comparable to that of retrieval of single-item information. However, the time course of temporal order judgments is successively slower for probes with items experienced less recently, consistent with a sequential recovery of less recent items. Order judgments induce a more effortful, time-consuming, and serial process of retrieval (McElree and Dosher, 1993) (Figure 4).

Another form of relational, or order, working memory task that has been extensively studied is the n-back task (Figure 2(d)), which is a continuous performance task in which items repeated at a particular delay are detected. In a 1-back condition, immediate repetitions are detected, while in a 2-back condition, repetitions with one intervening item are detected (Cohen et al., 1997; McElree, 2001). In related tasks, a stream of successive items is inspected for a particular ordered sequence, such as ‘472’. Like other order judgment tasks, the n-back tasks require the maintenance, retrieval, and processing of order information as well as rehearsal and information manipulation to continuously update the task-relevant working memory set. The processing demands in these tasks have also been associated with attention (Engle et al., 1999). The n-back order tasks show load-dependent activity in functional brain activation, as studied by functional magnetic resonance imaging (fMRI) (Cohen et al., 1997), reflecting these higher retrieval and rehearsal demands.

In contrast to item working memory tasks, even the simplest relational or order tasks place increased demands on retrieval processes, with relational judgments often reflecting serial processing. More demanding forms of the relational or order tasks, such as the n-back tasks, also require extensive manipulation of the memory set ('relational plus' tasks), or a high demand on executive function.

Complex Working Memory Tasks

Complex working memory tasks involve sequences of responses. The complex tasks, like relational or order tasks, emphasize temporal order or spatial
Figure 4. A relational task (judgment of recency) (Figure 2(c)) requires slow and serial processing. The task is to decide which of two item probes is the most recent in the working memory load of six items. Accuracy ($d'$) is expressed as a function of processing time for conditions grouped by the most recent item. The time course of relational judgment depends on the recency of the most recent item, reflecting a serial retrieval process. The upper panel shows fast retrieval for all tests involving the last (sixth) item (1v6, 2v6, etc.). The other panels show successively slower retrieval for tests in which the second to last (fifth) item was most recent, and the middle (third) item was most recent (data from McElree and Dasher, 1993). Reproduced with permission of B. Dasher.

order of the memory set in the response sequence. Among the most classical working memory tasks is the span task (Figure 2(e)). In this task, a working memory load of digits, letters, or words is presented, usually in temporal sequence, and is then recalled under instructions to reproduce the list exactly (serial order recall). The more demanding form, known as 'running span', interleaves processing distraction with the input of the working memory set (Figure 2(f)).

As was mentioned earlier, the modular view of working memory was developed and tested using the ordered word span task. One key observation was that the time taken to pronounce words determined the number of words that were successfully recalled in the correct order (the word-length effect) (Baddeley et al., 1975). Based on these observations, the duration of working memory was estimated to be approximately 1.5–2 seconds—the time required to pronounce a list of span length (that length of list which is perfectly recalled on average half the time). The idea is that if (subvocal) rehearsal can be completed before item traces are lost from working memory, then the list could be perfectly maintained (Schweickert and Boruff, 1986). One problem with this account is that the actual time taken to recall a span-length list may be 5–8 seconds, which is two to three times the estimate of 1.5–2 seconds (Dosher and Ma, 1998). Either the verbal memory trace is maintained in a phonological loop over intervals of up to 8 seconds, or complex attentional strategies maintain items in working memory during the extended act of recall. In common with the relational or order judgment tasks, the simple span tasks require the maintenance and retrieval of order information in working memory. In addition, these complex tasks require the coordination of a succession of recall products, and may involve complex rehearsal and information management.

Complex memory span (Figure 2(f)) makes the same demands as simple serial ordered recall, but each item in the working memory set is separated by a distracting task (Daneman and Carpenter, 1980). Perhaps because these tasks make extremely high demands on mental manipulation and transformation of information, they have been especially important in demonstrating the interrelationship between working memory capacity or efficiency in individuals and the general cognitive functioning of those individuals (Engle et al., 1999).

MODELS OF SERIAL MEMORY

Performance in serial memory tasks is characterized by a large and systematic set of phenomena. This does not just include the word-length effect (in which fewer longer words than shorter words can be remembered) and the reduction of successful recall under articulatory suppression (repetition of a distracting phrase) discussed in relation to the modular working memory framework (Baddeley, 1986). Additional phenomena include better memory for spoken than for written presentation, error rates that increase with serial position, and the dominance of transposition (order inversion) errors, as well as many others.
A number of precise computational models have been developed to account for performance in serial memory tasks. Different models are based, in some cases, on quite different principles and assumptions about representation, the causes of information loss, and information retrieval. Three examples of specific mechanistic accounts of working memory in sequential recall tasks are the feature model of Neath and Nairne (Neath, 1998), the primacy model of Page and Norris (1998), and the distributed associative memory model of Lewandowsky and Murdock (1989). This list is by no means exhaustive, but it is selected to illustrate the range of possible representations and processes in models of working memory. Future research must link these representational and process assumptions to the consequences of brain function and coding (for a network example, see O'Reilly et al., 1999).

In the feature model, items are represented by collections of modality-independent and modality-dependent features, with the items that are represented stored in order of input as a linked list. As each new item is stored, it over-writes similar features of the immediately previous item or items. In ordered recall, each item is recovered in order from an incomplete or noisy representation. Spoken presentation leads to the storage of a larger number of modality-specific features than visual presentation. Auditory similarity reduces memory due to over-writing of similar auditory features, while articulatory suppression over-writes verbal modality-specific features. Word length has an effect on performance because long words are stored in a sequence of syllabic structures. In contrast, in the primacy model, order is recoded as strength—items are stored in memory with higher activation strength for earlier list items. These strengths then undergo forgetting, or loss. At the time of recall, items are produced in the (noisy) order of strength. Rehearsal resets strength. Longer words take longer to produce or rehearse and experience more forgetting. Finally, in the associative model, individual items and associations between temporally adjacent items are stored in a single composite memory. Items, associations, and the composite memory are represented as a vector of feature values. Recall occurs by recovering successive items from the composite trace through chaining of recovered representations. Similarity between items may increase the error rates due to confusion between similar traces. Although each of the models fails to account for some aspect of the data, each of them also provides a good account of many of the phenomena in working memory that are exhibited in serial recall tasks.

As is evident, each of these models assumes a quite different form of representation and makes quite different assumptions about the nature of memory. Yet each of the models incorporates a mechanism for sequential readout of items from memory, and a mechanism of over-writing or interference to account for forgetting. As these and other quantitative models are developed further, they should provide an increasingly accurate explanation of the maintenance functions of working memory. Precise structural and functional descriptions will complement information about biological implementation.

**BIOLOGICAL SUBSTRATES OF WORKING MEMORY**

Working memory functions are now thought to reflect activity in a network of brain regions, including regions of the prefrontal cortex and more posterior regions in association cortex (Petrides, 1994). Important brain regions have been identified using a range of methods. Cellular recording of activity in non-human primates during various delayed response tasks has identified the continued activation during the delay (retention) period in certain prefrontal areas with storage of that information (Fuster, 1973). The time course of the delay activity has been shown to depend upon both response accuracy and the duration of the retention interval. Although initial demonstrations demanded spatial memory in responses, delay-period activity in the prefrontal cortex has also been demonstrated in a variety of nonspatial working memory tasks.

Regions of the prefrontal cortex have also been identified in a variety of brain imaging studies in humans (Cohen et al., 1997). However, the majority of these studies involved more complex forms of working memory tasks, such as the n-back tasks, which make extensive demands upon manipulation and transformation of the information to be remembered—executive function or control architectures. Some researchers have argued that activation in the dorsolateral prefrontal cortex reflects manipulation and transformation, rather than storage or retention (Owen et al., 1998). Some evidence from early lesion studies of patients with specific disorders of working memory and also from imaging studies has identified regions of posterior parietal cortex with simple registration and storage. The registration, storage, retrieval, and manipulation functions have yet to be segregated in analyses of working memory function as measured by brain imaging in humans. This will require the
measurement of brain activation with reasonable temporal resolution. In particular, this will necessitate the examination of tasks with clearly temporally segregated retention periods, and the comparison of different tasks with different demands for information manipulation. It will also require the measurement of brain activity in each of the major classes of working memory tasks (Figure 2). Critical new evidence segregating the activation patterns of these brain regions during specific time intervals, isolating storage, manipulation and retrieval in each of these task types, should be available within the next few years with improved methods and technology.

WORKING MEMORY, SHORT-TERM MEMORY, ATTENTION, AND EXECUTIVE FUNCTION

Understanding the functions and interrelationships between modules in working memory is necessary to a full understanding of the brain mechanisms involved, and to the construction of a complete information-processing model. Some theorists (Cowan, 1995) define working memory as the functional interrelationship between short-term memory, processes of attention that reactivate the short-term memory set, and executive function related to transformation, manipulation, and strategy selection. A substantitive task analysis is necessary for the segregation of these functions in behavioral analysis as well as in the analysis of brain activation. Future research instantiating working memory tasks within models of perceptual motor limitations, memory stores, and executive function (O’Reilly et al., 1999) may provide a precise theoretical structure for distinguishing these subfunctions of working memory.

WORKING MEMORY AND GENERAL COGNITIVE PERFORMANCE

The importance of working memory in general cognitive function is directly supported by the relationship between measures of working memory and performance in a variety of cognitive tasks (Baddeley, 1986). Measures of working memory retrieval time (Figure 3) have been correlated with general cognitive indices such as aptitude scores (Engle et al., 1999), and have been shown to differ for different developmental and other populations (Sternberg, 1975). The simpler modules of working memory, such as the phonological loop, appear to play a vital and quite specific role in developmental functions such as language learning (Baddeley et al., 1998), where the ability to repeat and maintain a phonological representation may be especially important in learning new words during development, or in learning new languages when supportive semantic information may be unavailable. Performance on working memory tasks, especially complex working memory tasks such as running span, with high demands on executive function (Figure 2(f)), is also correlated with performance on a wide range of other tasks, such as reading or problem solving. Thus, executive functions of working memory may be especially relevant for general cognitive function.

Working memory serves a basic human intellectual function. Elucidating the behavioral function of working memory, the brain activity that supports working memory function, and its relationship to general cognitive function represents a central component of understanding intelligent human activity.

References


**Further Reading**


NOTICE - Warning Concerning Copyright Restrictions.

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of Copyright Law.
Subject: DDS Article request from George Sperling
From: lrhaines@uci.edu
Date: Fri, 27 May 2005 15:33:31 -0700 (PDT)
To: dds@uci.edu

Delivery format: Web
Author: Lynn Nadel, Editor
Article Title: The Encyclopedia of Cognitive Science, Volume 4
Journal or Newspaper Title: The Encyclopedia of Cognitive Science
Date: 2003
Volume: 4
Issue: 569 - 573
Pages: 8
Call Number: BF 311 .E 53 2003 V. 4
Special Instructions: I need the chapter containing Fig 2 on page 571
Recharge Number: 9-406716-19900-3
Project Code: 
Department: CogSci
Send Request to: Langson Library
Name: George Sperling
Library Card Number: 21970001709327
Email Address: lrhaines@uci.edu
Affiliation: Faculty