

## Review Article

# Theories of Working Memory: Differences in Definition, Degree of Modularity, Role of Attention, and Purpose

Eryn J. Adams,<sup>a</sup> Anh T. Nguyen,<sup>a</sup> and Nelson Cowan<sup>a</sup>

**Purpose:** The purpose of this article is to review and discuss theories of working memory with special attention to their relevance to language processing.

**Method:** We begin with an overview of the concept of working memory itself and review some of the major theories. Then, we show how theories of working memory can be organized according to their stances on 3 major issues that distinguish them: modularity (on a continuum from domain-general to very modular), attention (on a continuum from automatic to completely attention demanding), and purpose

(on a continuum from idiographic, or concerned with individual differences, to nomothetic, or concerned with group norms). We examine recent research that has a bearing on these distinctions.

**Results:** Our review shows important differences between working memory theories that can be described according to positions on the 3 continua just noted.

**Conclusion:** Once properly understood, working memory theories, methods, and data can serve as quite useful tools for language research.

Working memory can be described as a limited amount of information that can be temporarily maintained in an accessible state, making it useful for many cognitive tasks. It is one of the most influential topics discussed in psychological science. One of the reasons for its popularity is the vast variety of activities and cognitive processes in which working memory is thought to play a role. As a real-world type of example, suppose a teacher tells the class that Earth is the third planet from the sun and asks a particular student to find it on a map of the solar system posted on a wall. The child must remember the first part of the teacher's speech (about the Earth's location) while processing the second part (the request for the child to find it on the map; cf. A. Baddeley, 2003). At this point, thoughts about performing in front of the class and how to handle that social demand may preoccupy working memory, competing with the assigned task. The point that Earth is the third planet must be retained in a ready form while the child implements a potentially tricky

routine of counting, starting not with the sun itself but with the planet closest to it. The child also has to remember to stop counting at the correct planet when the number 3 is reached and then, perhaps, look toward the teacher for feedback. The limits of working memory are such that there are many points at which this hybrid process can go awry because multiple skills compete for a limited working memory capacity. In a different kind of example, a young child can understand what is meant by a tiger only by holding in mind and combining three features, *big*, *cat*, and *striped*; a tiger is a big cat with stripes. These features distinguish a tiger from, in turn, a house cat (not big), a zebra (not a cat), and a lion (not striped; cf. Halford, Cowan, & Andrews, 2007).

## Definitions and Conceptions of Working Memory: A Brief Overview

Although examples like the ones presented above give us an idea of how working memory functions, it is often difficult to find one definition that encompasses all applications of working memory. Often, different theories—of working memory or otherwise—cannot be compared directly because the theories, though nominally on the same topic, actually are based on subtly different definitions of what is being studied. Cowan (2017a) examined the definitions of working memory commonly stated or implied in the research literature and listed nine definitions. Here, we cover only a

<sup>a</sup>Department of Psychological Sciences, University of Missouri, Columbia  
Correspondence to Nelson Cowan: cowann@missouri.edu

Editor-in-Chief: Sean Redmond  
Editor: Ron Gillam

Received October 13, 2017  
Revision received January 25, 2018  
Accepted February 18, 2018

[https://doi.org/10.1044/2018\\_LSHSS-17-0114](https://doi.org/10.1044/2018_LSHSS-17-0114)

**Publisher Note:** This article is part of the Clinical Forum: Working Memory in School-Age Children.

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

definition that should apply to all of the theories of interest and, then, more specific definitions tied to the major theories that will be described in detail.

In a definition that seems most generic and usable across different theories (Cowan, 2017a), working memory is a system of components that holds a limited amount of information temporarily in a heightened state of availability for use in ongoing processing. The definition does not depend on statements about the exact organization of components that may store or process information. This definition allows us to think of working memory information as separate from the rest of memory and uniquely important in carrying out cognitive tasks, and we believe that the field as a whole would not strongly object to this working definition.

To our knowledge, the earliest mention of the term *working memory* originated not from the study of the human brain but from the study of the computer. Computer scientists utilized the term *working memory* to refer to structures they set up within their programs to hold information that was needed only temporarily in executing procedures, such as solving geometry proofs (Newell & Simon, 1956). Although humans are unable to manage multiple temporary storage structures at once like computers, still, it is instructive to realize that the need for temporary storage arose in the process of inventing problem-solving routines. The use of the term *working memory* for human research started with Miller, Galanter, and Pribram (1960). They considered working memory as a part of the mind that allows us to operate successfully in life, completing our goals and subgoals by storing the useful information needed to execute these planned actions. For example, the goal of furthering one's career can have a subgoal of getting an academic degree, with a sub-subgoal of making it to class today, a sub-sub-subgoal of getting dressed, and so on, down to one's momentary activities. Forgetting information at the wrong time leads to errors.

A. D. Baddeley and Hitch (1974) jump-started the field of working memory, and they defined the state of affairs preceding their paper as the short-term or immediate memory view on the basis of what they called the *modal model* or very usual type of model at the time. The most-often-cited example was the work of Atkinson and Shiffrin (1968). In that work, short-term memory was represented by a single mechanism that temporarily held information to be used in processing. The most common task leading to that conception was a simple span task in which, on each trial, a list of verbal items was presented and was to be repeated back verbatim; the longest list that could be repeated correctly is the memory span. Atkinson and Shiffrin focused also on control processes used to shuttle information between stores, as when knowledge is used to enrich the contents of the short-term store.

In the research-rich book chapter of A. D. Baddeley and Hitch (1974), the term *working memory* came to them as they attempted to distinguish their views from the modal model. Their definition of working memory was as a multi-component system to store temporarily information as it is processed. Baddeley and Hitch found results that they could

not represent by a single process, as if they had to break the box representation into multiple boxes, which they called multiple components of a system they termed *working memory*. One component held verbal information (the phonological store), another component held visual and spatial information (the visuospatial store), and yet another component was a processor (the central executive), responsible for moving information into the stores and using them to guide behavior. In the most recent version of A. Baddeley's (2000) model, another component (the episodic buffer) temporarily holds semantic information and associations between different kinds of information (e.g., face-to-name links).

In contrast to simple span tasks, the tasks that A. D. Baddeley and Hitch (1974) presented typically involved retaining a list in memory while carrying out another process, like completing a reasoning problem, and then recalling the list. When multiple stimuli have to be processed, there is supposed to be interference between stimuli that are being retained or processed using the same kinds of information codes, such as two verbal tasks or two spatial tasks, but not interference between information held in different codes, such as a verbal list to be recalled and a concurrent spatial task. Interference is supposed to occur only when working memory representations of two or more stimuli depend on the same component or store at the same time.

Many researchers interested in the application of working memory to real-world types of cognitive function, including language processing (e.g., Daneman & Carpenter, 1980; M. A. Just, Carpenter, & Woolley, 1982), have adopted a slightly different emphasis on the basis of the work of A. D. Baddeley and Hitch (1974) and follow-up work (e.g., A. Baddeley, 2000). They distinguish between the situation when one only has to store and then repeat information without processing or manipulating it, which they call *short-term storage*, and the situation in which one has to manipulate the stored information, which they term *working memory*. For example, if you hear a list of grocery items and just have to repeat the list, that would be termed *a test of short-term memory*, whereas if you hear a list of grocery items and have to repeat them in a different order, with vegetables and fruits first, dairy items second, and other items third, that would be termed *a test of working memory* (though others use the terms slightly differently; see Cowan, 2017a). These researchers were not so concerned about whether this working memory was a multicomponent system or not.

## Organization

We will next discuss some ways in which working memory is important for language. Then, we will present three often-discussed theories that illustrate different ways in which working memory can be conceived (the already-mentioned theories of Atkinson & Shiffrin, 1968, and A. Baddeley, 2000, and a different conception by Cowan, 1988). Finally, we will discuss working memory theories within an organizing framework in which we point out

three dimensions on which the theories differ, namely, (a) the degree of modularity, (b) the degree of reliance on attention, and (c) the purpose of the theory as elucidating individual differences versus group means. These dimensions will be presented as continua on which different theories can be placed. In a final section, recent research on working memory pertaining to these dimensions will be highlighted. The evidence suggests a fortunate convergence of the different theories in recent years, and implications for future language research are discussed.

## The Importance of Working Memory in Language Processing

The other articles in this issue of the journal provide a detailed picture of the use of working memory in language, so here, we simply give an initial glimpse of this use to illustrate the relevance of our descriptions of models of working memory. In academically relevant areas, including problem solving, learning, reasoning, and mathematics (numerical, symbolic, and spatial), among other areas, working memory capability has often turned out to be one of the best predictors of cognitive performance. For our purposes, we will briefly discuss how working memory is important to language comprehension and production.

Although materials of all sorts can be held in working memory, it has long been noticed that different materials are not on equal footing. Conrad (1964) found that even when letters were presented in visual form to be remembered, mistakes consisted primarily of acoustic rather than visual confusions, suggesting that participants were in some way converting visual materials to a phonological (speech-based) code. Subsequent work on the effects of manipulations to encourage or discourage the use of speech codes suggested that the special privilege of verbal materials is that they can be covertly or overtly repeated, or rehearsed, without much effort to keep the working memory active. This kind of concept about the role of language was represented in the A. D. Baddeley and Hitch (1974) theory of working memory. The relation to language was amplified when it was determined that the ability to remember and repeat phonological sequences, such as multisyllabic nonsense words or short series of words, was critically important for vocabulary learning (e.g., A. D. Baddeley, Gathercole, & Papagno, 1998).

The focus of A. D. Baddeley and Hitch (1974) on phonological processes and rehearsal was important in order to make intensive progress in understanding one part of the working memory system and how it actually operates. Other researchers were interested in working memory and language on a more holistic level in order to determine how working memory functions for a common task, such as reading. Daneman and Carpenter (1980) and Case, Kurlund, and Goldberg (1982), therefore, devised working memory span tasks in which multiple components are presumably involved. In a reading span task, Daneman and Carpenter presented lists of sentences for which the participant had to do a comprehension task (engaging processing) while also

remembering the final word of each sentence (engaging storage). After the last sentence, the list of sentence-final words was to be recalled. Performance was assessed as the number of sentences that could be processed correctly while still permitting correct recall of the final words of the sentences. Case et al. similarly devised a counting span task in which series of arrays of simple objects were to be counted and the sum of each array was to be retained in memory and, then, recalled after the last array was counted. These complex span tasks correlated much better than simple digit span with verbal abilities, including reading (Daneman & Carpenter, 1980), though it was later observed that complex span tasks also correlate well with aptitudes across domains, not just language aptitudes (e.g., Cowan et al., 2005; Kane et al., 2004).

A great deal of the research on the implications of working memory on language processes originates from research on language disorders (e.g., de Jong, 1998; Gathercole & Alloway, 2006; Swanson, 1999). Gathercole and Baddeley (1990) studied children with developmental language disorders compared with control groups on multiple working memory-related tasks. Their results showed that children with language disorders performed at lower levels than age-matched peers on nonword repetition tasks and, sometimes, even lower than younger peers matched on vocabulary and reading. Another experiment in the study showed that children with language disorders did not differ from peers on their ability to rehearse information. These and other supporting findings suggested that children with language disorders do have working memory storage deficits, which could contribute to, or perhaps even cause, the disorders. Subsequent research goes further to try to understand the mechanisms of working memory deficits and language disorders, including specific language impairment (see Marton & Schwartz, 2003; Montgomery, 2003; Weismer, Evans, & Hesketh, 1999). Other research shows how working memory deficits in retention of serial order information are involved in language impairment (Gillam, Cowan, & Day, 1995) and dyslexia (Cowan et al., 2017; Majerus & Cowan, 2016).

One growing line of research deals with the implications of working memory in second language acquisition and use. In a world where many individuals are exposed to and juggle more than one distinct language, understanding the processes that underlie successful processing is of utmost importance. Working memory is thought to be a critical ability in the acquisition of a second language, though the mechanisms remain unclear (Cowan, 2015). In an example of expert language use, cross-language interpreters face the task of trying to hold the information spoken by the original speaker and what they have already translated, as well as the gist or the topic of the conversation at hand (Cowan, 2000/2001). Their work requires intensive attentional filtering and attention switching, as well as temporary storage, or working memory capacity.

Although there is much supporting evidence for the importance of working memory in language processing, the exact role of such a source has been debated in several

ways in the last few decades. One such line of debate concerns the role of working memory in syntactic processing. M. Just and Carpenter (1992) proposed a theory in which language comprehension is constrained by working memory capacity. Included in this theory was a proposal that the modularity of language processing is best explained as a capacity constraint rather than one of architecture. Thus, individuals with smaller working memory capacities may not have enough available activation to process and store nonsyntactic information during syntactic processing. Individuals with larger working memory capacities should then be able to handle both syntactic and nonsyntactic information at once and may experience an influence of the nonsyntactic information on syntactic comprehension. These differences might cause some people to appear to have more modular language processing than others, but the authors proposed that it all depends on their working memory capacity for language, not a distinct separation of modules.

M. Just and Carpenter (1992) called upon a previous study (Ferreira & Clifton, 1986) in which readers processed garden-path sentences with or without semantic information that could steer the interpretation of syntax. In the sentence, “The defendant examined by the lawyer turned out to be unreliable,” it is at first possible to think that the defendant is the one doing the examining, the garden-path interpretation that leads participants to spend a long time looking at the word *by*, presumably because their initial interpretation was wrong. In the sentence, “The evidence examined by the lawyer turned out to be unreliable,” in contrast, the nonanimacy of the subject “evidence” should be a clue that the agent who does the examining comes later in the sentence; yet, participants still dwelled on the word *by*, showing that they were captured by the garden-path interpretation even though it is semantically implausible. Just and Carpenter replicated the study, this time separating individuals according to their span. Low-span individuals were still led down the garden path, as previously found, whereas high-span individuals were able to take into account the nonsyntactic information. The authors concluded that syntactic processing in high-span individuals was not modular but interactive, suggesting a domain-general capacity that applied to both syntax and nonsyntactic contextual information. Recent evidence also suggests that high-span adults are more likely to keep their options open longer when trying to resolve the meaning of ambiguous printed sentences; lower span adults tend to break up the text up into smaller chunks and seize upon convenient interpretations on the basis of the chunks without waiting for more input (Swets, Desmet, Hambrick, & Ferreira, 2007).

In a critique of some aspects of the capacity-based theory, Waters and Caplan (1996) proposed that Just and Carpenter’s interpretation of the garden-path results was not adequate. They noted that their method was not actually a direct replication of the original methods utilized by Ferreira and Clifton and, therefore, could not be interpreted in the same ways. Also, they pointed out that the data reported by Just and Carpenter still showed that individuals with both low and high spans experienced the

garden-path error for some sentences. Waters and Caplan suggested that these trends in the data only further confirm the modularity view of syntactic processing. The authors also argued that, if the Just and Carpenter theory is correct, language processing results should show differences in overall sentence processing that are related to working memory capacity. They note that this difference was not always found in some previous studies and that, in one study, low-span individuals were able to use pragmatic information to help assess sentence meaning but high-span individuals were not (King & Just, 1991).

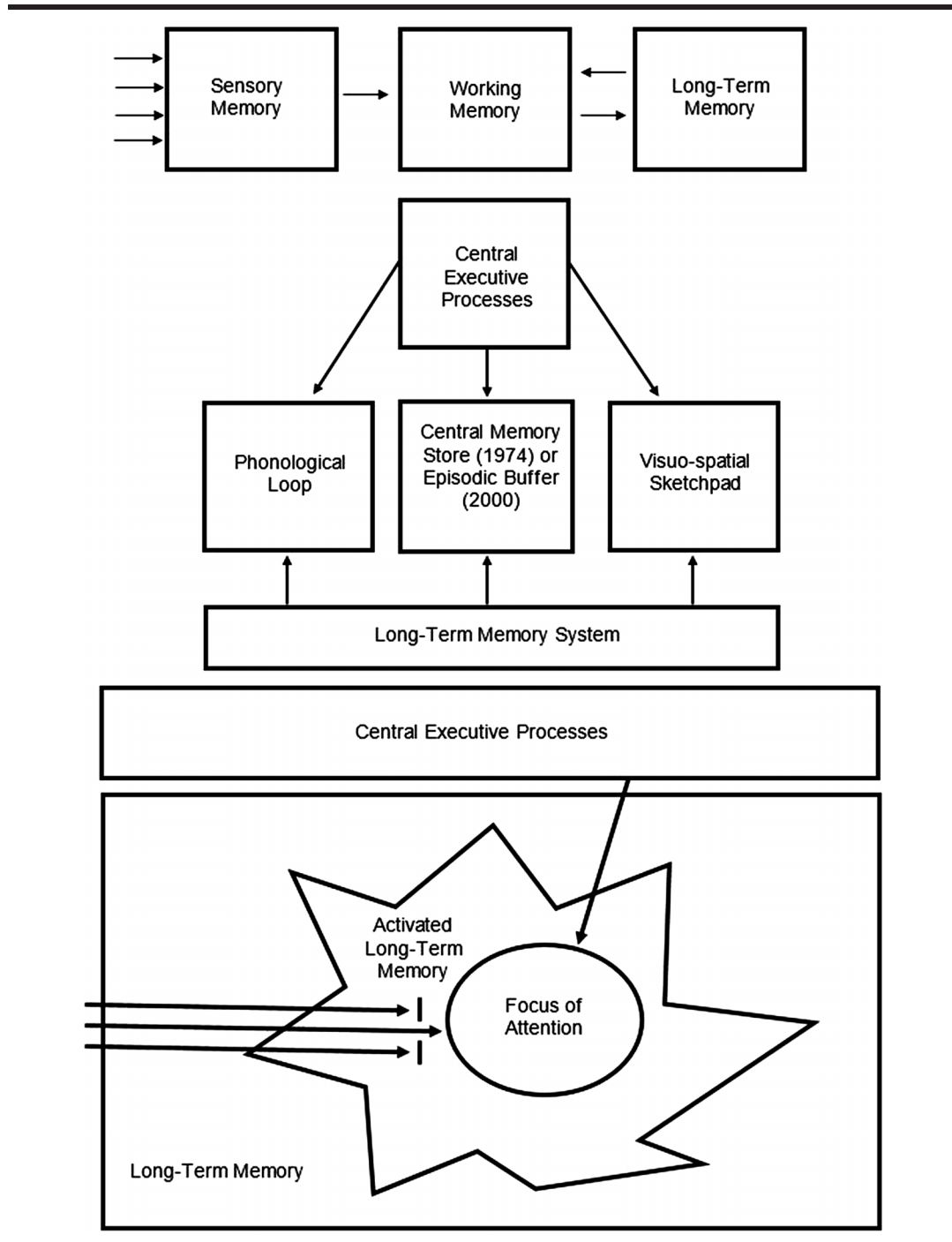
Though Just and Carpenter disagree with Caplan and Waters on the modularity of language processing and on the role of working memory during this processing, one aspect of their theories that they share is the proposal that linguistic knowledge and working memory are two separate entities. Carpenter, Miyake, and Just (1994) offered evidence from readers with brain injury or disease in which the lexicon and production rules remained intact but storage and processing of language were severely impaired. They proposed that these results supported the idea that what one knows about language (i.e., language competence) and how language is processed (i.e., language performance) are two different entities. However, MacDonald and Christiansen (2002) proposed a resolution in which knowledge and capacity actually cannot be considered separately because the processing and storage stems from a passing of activation through a common learning network.

In sum, though some of the exact mechanisms of the involvement of working memory in language processing have been debated and are uncertain, there is plenty of evidence supporting a conjoining of the two fields of study. Working memory is an important cognitive skill to consider when approaching the study of individual differences in language processing, comprehension, and production, as well as language development and disorders.

### Three Examples of Working Memory Theories

To explore in greater detail some theories of working memory, Figure 1 illustrates three often-mentioned theories. The top panel shows a schematic depiction of what Alan Baddeley has often light heartedly called the *modal model*, meaning the type of model of which the most instances existed (circa the late 1960s). The best-known example is that of Atkinson and Shiffrin (1968), though a precursor is found in a footnote of a book by Broadbent (1958). A large amount of incoming sensory information is mostly forgotten, but a small amount of the information advances to a working memory, where it is enhanced by long-term memory information and temporarily retained. Working memory is also the basis of the formation of new long-term memories. As evidence of the need for separate short-term and long-term mechanisms, Atkinson and Shiffrin stressed the effects of hippocampal lesions, which show diminished long-term memory storage with preserved short-term storage (e.g., Milner, 1968). Their model also placed an emphasis on control processes (not shown), which strategically help to

**Figure 1.** A depiction of three models of working memory. Top: what Baddeley has termed *the modal model*, for example, after Atkinson and Shiffrin (1968); middle: a version of the A. D. Baddeley and Hitch (1974) model or a recent revision of it by A. Baddeley (2000); bottom: embedded-processes model of Cowan (1988, 1999).



recirculate information in working memory and shuttle information between working memory and long-term memory.

The middle panel of Figure 1 shows the model that sparked the field of working memory, initiated on the basis of a large number of experiments (A. D. Baddeley & Hitch, 1974) and then put through several iterations (A. D. Baddeley,

1986; A. Baddeley, 2000). The key difference between this and the modal model is that working memory here has been split into a few different specialized stores and a more general store. One specialized store (left-hand box in the middle panel) is for phonological information, and another (right-hand box in the middle panel) is for visuospatial information.

The more general store (shown between the phonological and visuospatial stores), called the *episodic buffer*, is not specialized for any one kind of information but available to link different kinds and is possibly tied to attention. Long-term memory feeds category information into the stores used to guide the interpretation of sensory input. Similar to the modal model, Baddeley's model includes some set of mechanisms, collectively called the *central executive*, that govern the strategic control of information. This component may be even more sophisticated than the control processes of the modal model because, in Baddeley's model, there are more separate stores to contend with and, therefore, more potential mnemonic strategies and manners of processing information. Among its other activities to schedule and prioritize information transfers and behaviors, the central executive initiates a rehearsal process to prevent decay of information from the stores.

The bottom panel of the figure depicts the embedded-processes model proposed by Cowan (1988), named by Cowan (1999), and enhanced with a clearer notion of its central, capacity-limited portion by Cowan (2001). Unlike the Baddeley model, which was focused around the kinds of effects he and his colleagues were observing in the laboratory, Cowan's model was an attempt to establish a more general framework for information processing insofar as it was known. Information comes in from the environment through a very brief sensory store (as depicted by rightward-pointing arrows), activating features in long-term memory corresponding to the sensory properties of the incoming information and its coding: phonological, orthographic, visual, and other simple features from the senses. The phonological and visuospatial stores are not separated in this model because it is assumed that there is a rather complex taxonomy and that it is uncertain which stores are basic, which are overlapping, and so on. In place of showing separate stores, the same evidence is accommodated by the simple proposal that new input overwrites or interferes with previous activated information with similar features. As in Baddeley's model, the information supposedly decays if not rehearsed or, alternatively, is more quickly and nonphonologically refreshed via attention (Barrouillet, Bernardin, & Camos, 2004; Cowan, 1992; Raye, Johnson, Mitchell, Greene, & Johnson, 2007).

Some kind of filtering function that limits how much information gets into working memory seems necessary in any model of processing (cf. Broadbent, 1958). Cowan (1988) suggested a specific mechanism for it, dishabituation of orienting. In the orienting response, an individual's attention is turned to a stimulus that stands out from the background in the environment. It may be a sudden change in the environment or a newly presented item of special meaning to the individual. With repetition, the novelty soon wears off, and the orienting response dies down or habituates. In such a mechanism, all information from the environment stimulates physical features, but a neural model of the environment is built up over time, and only the information discrepant with the model causes dishabituation or a restrengthening of a previously weakened response and,

thus, attracts the focus of attention. That focus also can be directed by the central executive, which allows it to pick up more abstract, semantic information voluntarily. The focus of attention allows a coherent organization and interpretation of the information it contains, but that information is limited to a few separate, known items at a time. The separate items can be linked to form a new memory that becomes part of the long-term record. When items leave the focus of attention, they still remain activated for a while. These previously attended, meaningful items, along with the never-attended physical features of the rest of the environment, all contribute to the neural model, and any noticed change from the neural model attracts attention. The changes can be physical, often regardless of attention, or semantic, usually with attention. Thus, the activated features from long-term memory, including any newly formed memories, along with the current focus of attention, together comprise the working memory system. This system is limited by interference and decay of activated memory and by a capacity limit of the focus of attention. Fatigue of the focus of attention also is possible.

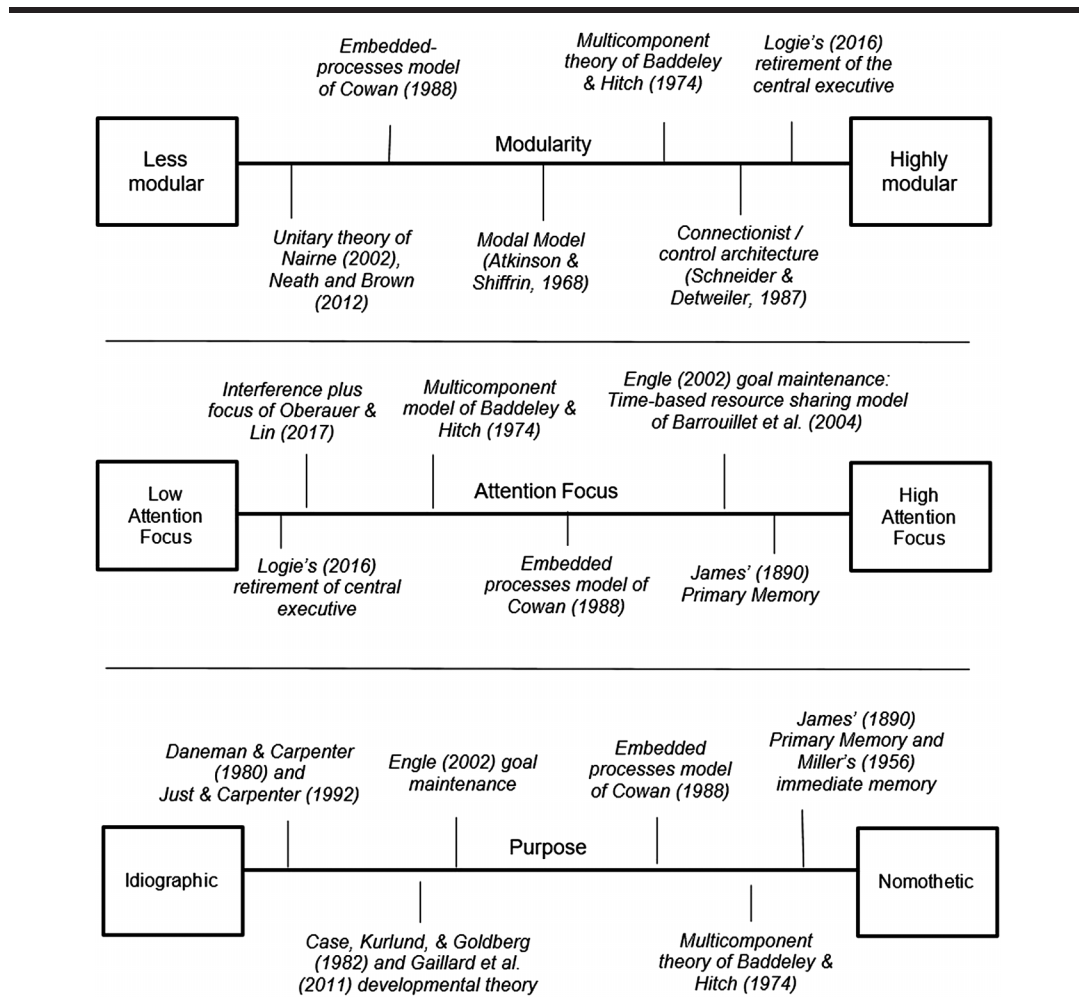
## Theories of Working Memory Distinguished on Several Continua

In the next part of this review, we will differentiate some well-known and representative theories of working memory, beyond those we have discussed in detail, by focusing on three main continua that tend to differentiate them: the degree of modularity, the role of attention, and the nomothetic versus idiographic purpose. Though these continua are not the sole discriminating issues, they provide a useful orientation for understanding differences among working memory theories. We will name theories that lie on either extreme of each continuum and also theories that tend to straddle the middle, at least as we perceive them. Other theories, in addition to the ones previously described, will be mentioned briefly within the continua to assist further exploration. We will also highlight language, speech, or auditory research that supports or rebuts relevant theories. Figure 2 illustrates the continua and how we have rated various theories on them.

### *Degree of Modularity*

Modularity deals with the organization of the system of working memory and how compartmentalized it is. If working memory were a house, a highly modular theory would be a house with many rooms, or modules, each designated to a specific type of information. A less modular theory would have fewer, bigger rooms that process and store all types of information. Thus, modules of working memory are functioning parts of the system that store, maintain, or process different types of information independent of one another. Information can be categorized on the basis of different types of characteristics. Some theories that could be considered modular (to a degree) separate stores on the basis of the amount of time the information

**Figure 2.** Three continua discriminating models of working memory from one another. Top, the degree of modularity; middle, the degree to which attention underlies the storage and handling of stored information in the model; bottom, the dependence and use of idiographic (individual difference) versus nomothetic (group mean) information.



has been held (short term vs. long term). Other, more modular theories may take time into account but also separate stores on the basis of the type of information (verbal, visuospatial, etc.). The modules, however, are not necessarily separate brain areas and could overlap neurally. By analogy, the U.S. government in Washington, D.C., includes three branches (modules), but any one geographical area in Washington can include representations of two or even all three branches.

Certain consequences arise from regarding working memory as either modular or not. In a modular theory, if one module is at capacity in terms of the amount of information it can actively store or process, other modules are still available for use. Less modular theories imply instead that, when these nondiscriminatory areas of working memory are at capacity, no type of information beyond capacity will be processed or stored successfully. In what follows, we examine a theory with no modularity and, then, consider different types and degrees of modularity.

### Unitary Theories With No Working Memory/Long-Term Memory Distinction

If working memory is to differ from long-term memory, we can think of two basic ways in which this difference can occur. There must be information in working memory that is limited to a certain time period, a temporal decay property, or limited to a certain amount of information, an item capacity property. Either of these properties could be modulated by the amount of interference. Nevertheless, if they do not exist at all, there would be only one kind of memory as posited by unitary memory theories, which forgo any separation of short-term or working memory versus long-term memory. (We will see that some such theorists still exist.) One of the earliest researchers to propose such a view was McGeoch (1932), who sought to argue against Thorndike's (1914) proposed law of disuse. Thorndike suggested that, when a stimulus-response association is not activated for a long time, the strength of the connection decreases. One might then distinguish between short-term,

labile memories versus longer-term memories that remain because of repeated use. McGeoch argued, however, that disuse does not always mean forgetting. For example, he referred to a study showing recovery of conditioned responses during a period of inactivity following experimental extinction. If memories do not always grow weaker over time, the argument goes, there is no reason to talk of a short-term memory separate from long term, an argument that was reinforced by Underwood (1957). He proposed that most forgetting came from some combination of interference that was proactive (from previous stimuli in the experiment or in everyday life) and retroactive (from information received between the stimulus and test), both of which could impede fully accurate memory of target items. According to this view, the recency of a memory does not directly distinguish it from older memories; only the amount of interference that has occurred does.

Against unitary theory, Peterson and Peterson (1959) carried out a study in which letter trigrams were presented, and before they were to be recalled, a variable period of counting backward by 3 was introduced to prevent rehearsal. The researchers found that letter memory declined dramatically as the period of counting backward increased from very short to 18 s, despite the dissimilarity of letters to numbers. This decline was taken as an indication that a short-term memory of the letters decayed over time. Keppel and Underwood (1962), however, showed that, in this kind of procedure, the dramatic drop-off did not occur at all on every participant's first trial but developed over trials. They suggested that proactive interference from previous trials increases as the retention interval on the present trial increases, removing the temporal context of the most recent items. Keppel and Underwood's interpretation from the unitary memory view was that proactive interference alone accounts for the effect of the retention interval. An alternative, two-store interpretation that Keppel and Underwood did not consider is that there could be a short-term memory that decays over 18 s and also a long-term memory of the present memoranda that can be used, at all retention intervals, on the first few trials. Proactive interference builds up across trials quickly, and after it has built up, long-term memory no longer contributes much to recall; this change can explain why forgetting over retention intervals appears in later trials, as the participant becomes more dependent on a temporary short-term memory.

In another example of evidence seemingly favoring unitary memory theory, Bjork and Whitten (1974) challenged the notion that, in free recall of a verbal list, a pronounced advantage for recall of the most recently presented items (recency effect) is the evidence that those items are recalled from short-term memory. Glanzer and Cunitz (1966) had shown that requiring a distracting task of counting aloud for 30 s before written recall abolishes the recency effect, and they attributed that effect to the degradation of a short-term store. Bjork and Whitten, however, reinterpreted the recency effect in terms of the temporal distinctiveness of the end of the list or how separate in time from one another items on a list seem. Better temporal distinctiveness is

supposed to facilitate the task of retrieving the right information from memory. As the distraction period continues, loss of that distinctiveness occurs and, thus, increases proactive interference. By separating all list items by distracting tasks, they were able to preserve a recency effect despite a distraction-filled period after the list and before recall, presumably because that final period could no longer reduce temporal distinctiveness very much in these spaced lists.

Most current theorists acknowledge that there is sometimes a contribution of temporal distinctiveness and proactive interference, as the unitary theorists assume. However, they also point to evidence that a recency effect obtained with distractors between items has properties different from a recency effect obtained even when temporal distinctiveness is low, evidence for a separate short-term store after all (see discussion of the "monistic view" by Cowan, 1995; and see Davelaar, Goshen-Gottstein, Ashkenazi, Haarman, & Usher, 2005).

*Absence of decay in unitary theories.* One of the main issues that separates unitary theories of memory from theories that are more modular is that proponents of unitary memory theories do not believe that memory decays over time. Nairne (2002) suggested that certain memory cues (e.g., how pronounceable or tangible to-be-remembered items are) affect short-term retention just like they do long-term retention and that rehearsal and decay prove inadequate to explain forgetting. The original evidence for decay under cross-examination by Nairne was that individuals can recall lists of about as many items as they can repeat in about 2 s (A. D. Baddeley, Thomson, & Buchanan, 1975). The speed of repetition was assumed to approximate the speed of covert rehearsal and could be manipulated both by presenting words that took less or more time to say and by correlating performance with individual differences in the repetition rate. Nairne, however, pointed to a study by Schweickert, Guentert, and Hersberger (1990) showing that, when participants were presented with lists of similar and dissimilar words at the same pronunciation rate, there were still span differences between the two types, suggesting that time alone is not a sufficient account of forgetting. In general, Nairne argued that, although time is correlated with forgetting, it is the events that happen during a particular time period that are important for the loss of memory, not the passage of time. Therefore, he suggested that theorists should move on to a model of memory that recognizes short-term retention as largely cue driven. Evidence for cue-driven accounts of short-term retention includes characteristics of stimuli, such as lexicality, word frequency, or concreteness resulting in differences in recall. An even stronger statement against decay has been made (Neath & Brown, 2012) to the effect that only distinctiveness, interference, and retrieval context make a difference. Jalbert, Neath, and Surprenant (2011) found that, when short and long words were matched for neighborhood size (the number of words similar to the target word in linguistic features), the word length effect was eliminated. Oberauer and Lewandowsky (2008) showed, against the expectation on the basis of decay, that the passage of time during recall made little difference,



even with suppression of rehearsal and another, nonverbal task to engage attention.

### Theories Distinguishing Working Memory From Long-Term Memory Based on Decay of Items From Working Memory

A. D. Baddeley et al. (1975) and A. D. Baddeley (1986) invoked decay and rehearsal to explain why participants could recall lists of as many items as they could read aloud or recite from a memorized series in about 2 s. Presumably, the memory trace of the entire list had to be rehearsed in that amount of time or some of the items would be lost through decay. Barrouillet et al. (2004) proposed the same theory except that, in place of rehearsal, they proposed that refreshing through attention could be used. The evidence consisted of a negative, linear relation between the memory span and the proportion of time between items that was occupied by a distracting task, termed *the cognitive load*. The notion is that the cognitive load prevents refreshing and, therefore, allows items to decay. More recent work has suggested that either rehearsal, within the verbal domain, or attention-based refreshing, regardless of the type of materials, might be used together in various combinations (Camos, Mora, & Oberauer, 2011; Vergauwe, Barrouillet, & Camos, 2010).

These theories assume decay in the absence of rehearsal and refreshing and rely on that assumption but generally have not observed decay directly. Ricker, Spiegel, and Cowan (2014) did find that arrays of unfamiliar characters were forgotten over a number of seconds as a function of the retention interval between the array and a recognition probe; this trend was observed even when a temporal distinctiveness account could be ruled out. Subsequent work suggested that decay occurred only when there was not sufficient time after the presentation of stimuli for them to be well encoded into working memory in the first place (working memory consolidation: Ricker, 2015; Ricker & Cowan, 2014; Ricker & Hardman, 2017).

Although the decay observed by Ricker and colleagues allows one version of the embedded-processes model of Cowan (1988, 1999), it is problematic for some other theories. Ricker et al. actually have shown very little decay in situations similar to those in which decay has been used to help account for the 2-s rehearsal limit for lists that can be recalled (A. D. Baddeley et al., 1975) and for the cognitive load function (Barrouillet et al., 2004).

### Working Memory Distinguished From Long-Term Memory by Working Memory Capacity Limits

In place of decay, there could be an interference-based loss rate that depends on the amount of information held concurrently (Davelaar et al., 2005; Melton, 1963), that is, functionally, some sort of capacity limit. For example, in a Peterson and Peterson-type procedure (lists to be recalled after counting backward), the rate of forgetting is steeper when there are more letters in the set to be remembered, presumably because of within-set interference (Melton, 1963; cf. Murdock, 1961).

According to Cowan's (1988, 1999) embedded-processes model of working memory, the focus of attention is quite limited in capacity. Cowan (2001) explored what the average individual's memory span is when stimuli are presented in a way that prevents mnemonic strategies like rehearsal, chunking, and grouping. Chunking is the process of using what one already knows to make larger collections of items, reducing the amount to be remembered; an example is remembering the list IRSCIAFBI more easily as three acronyms for government agencies (the Internal Revenue Service, the Central Intelligence Agency, and the Federal Bureau of Investigation). *Grouping* refers to the process of combining items to form new collections that may be rapidly memorized. For example, one may memorize a list of nine digits by mentally separating the digits into groups of three (e.g., 674, 891, 532). When strategies such as these are prohibited, typical span for various kinds of materials (both verbal and nonverbal) is reduced from Miller's (1956)  $7 \pm 2$  to about  $4 \pm 1$ , on average (Cowan, 2001). The limit seems to hold for a wide variety of item types, though sometimes the observed capacity is lower because memory of complex items does not capture all of the details of the items (Awh, Barton, & Vogel, 2007).

There has been a challenge from theorists who believe that the observed capacity is actually a fluid resource that can be spread thinly over all items in an array or series (e.g., Ma, Husain, & Bays, 2014). However, recent work suggests that, if this is the case, after about three items, the fluid resource must become so thin as to be of no use (Adam, Vogel, & Awh, 2017), essentially removing empirical differences between the finite-slots and fluid-resources theories of working memory capacity limits.

### The Modularity Continuum

The top panel of Figure 2 shows a continuum of some models of working memory arranged from less modular on the left to highly modular on the right. The unitary theory is of course considered nonmodular. The embedded-processes model is just slightly more modular because its two mechanisms for working memory are nested rather than separate, with both of them nested within the long-term memory system. The modal model has separate short-term and long-term stores but still no proposed, specific structure within short-term (i.e., working) memory. The multi-component model is yet more modular, with separate stores for different types of code (verbal-phonological, visual-spatial, and semantic-binding).

We also include a couple more models in Figure 2 that are not discussed in detail. A model by Schneider and Detweiler (1987) actually goes in an even more modular direction, suggesting, at a microscopic scale, separate modules for auditory, speech, lexical, semantic, motor, mood, context, and visual stimuli, all under higher levels of control. In perhaps the most modular approach, Logie (2016) proposes that there are not only modules for specific kinds of materials, as in the multicomponent model, but also modular mechanisms replacing the central executive (cf. Vandierendonck, 2016).

Finally, note that, in the field of language, there similarly have been lively debates about whether language is represented in the brain in a very modular way (in which syntax is insulated from other aspects of language processing) or in a less modular way (in which syntax is one outcome of a general process limited by working memory constraints). It is possible that the more (or less) modular language theories naturally line up with the corresponding more (or less) modular working memory theories, and considering the nature of working memory and language modules together might shed light on the general nature of cognition, as well as yielding practical insights into the best ways to teach language and remediate language disorders.

### **Role of Attention**

It is generally, though not uniformly, the case that less modular theories of working memory have a higher reliance on attention. The main reason is that attention is conceived as the storage device that is limited but that can seize upon any kind of information, retaining, for example, some verbal items, some visual images (which may be related to the verbal information, as in a television commercial) and, even, some touches, musical sounds, and other sensations that have been meaningfully interpreted. Any such general storage across domains is capacity limited in that, although people perceive the entire scene (e.g., an arrangement of objects that looks like a kitchen), there is inattentional blindness for the exact properties of all but a few attended aspects of the scene. If the scene flickers or attention is drawn to a certain aspect of the scene, it is possible to replace one object with another, such as substituting a coffee maker with a toaster or with nothing at all, and observers tend not to notice except in rare instances in which attention was already focused on the changing object (e.g., Simons, 2000).

If working memory is limited by how much material is included in the focus of attention at once (Cowan et al., 2005), there are important implications for language processing. The easiest way to process language, much like processing visual materials, is to fit the received language input into a comfortable scheme that seems right without necessarily attending to all of the details. Results of Patson, Darowski, Moon, and Ferreira (2009) suggest that this is often the case. Adults who read a sentence like “While Janice dressed the baby slept” often came away with an impossible interpretation of that sentence (in this case, that Janice dressed the baby while the baby slept). Inattentional blindness to the part of the syntax would seem like a case for an attention-based working memory store that is indeed involved in ordinary language processing, regardless of language competence.

### **The Attention Continuum**

The middle panel of Figure 2 shows a continuum on the basis of the degree of usage of attention by working memory, from low (on the left) to high (on the right). Logie’s (2016) formulation seems not to subscribe to the notion of attention at all. Oberauer and Lin (2017) follow Oberauer’s

earlier work by subscribing to a single-item focus of attention in most situations, though the attention focus is capable of expansion when, say, two items need to be considered together. The multicomponent model makes rather more use of attention at least for processing, in the form of the central executive and its choices. The extent to which storage also relies on attention is a question currently in flux within that approach. In the embedded-processes model, attention is used not only for processing but also clearly for storage. Engle (2002) and Barrouillet et al. (2004) are similar in that one attention process seems critical for performance (correct goal maintenance in the face of interference and distraction, e.g., Kane & Engle, 2000; or refreshing of items before they decay). Finally, James (1890) discussed a mechanism that was nothing but the attention focus: primary memory that was essentially the same as the information in consciousness, most comparable to the focus of attention component of Cowan’s model.

In the discussions of language disorders, there have been considerable debates about the degree to which the disorders stem from automatic components of processing versus those that depend on attention and central executive function. Keeping in mind the alternative models of working memory that differ on the role of attention should help to inform this debate.

### **Nomothetic Versus Idiographic Purpose**

It is natural that some researchers are most interested in using working memory models to understand individual differences, known as *idiographic information*, whereas others are interested to understand how humans process information in general, known as *nomothetic information*. What might be less well-appreciated is how these approaches can actually work together. For example, if one wanted to distinguish between different modules or mechanisms, nomothetic researchers could hope to do so by showing dissociations within an individual (such as the findings of A. D. Baddeley & Hitch, 1974, indicating that a separate memory load did not reduce the recency effect in free recall, or that two sets of phonological materials interfere with one another more than one phonological set and one visual, nonverbal set). Sometimes, however, idiographic information is used for a similar purpose of model description, under the assumption that tests that assess a particular mechanism within the working memory system (e.g., the phonological loop) will yield individual differences that do not completely correspond to the individual differences observed in tests of a different mechanism (e.g., the visuospatial sketchpad). It was from this perspective that Gathercole, Pickering, Ambridge, and Wearing (2004) used structural equation modeling to show that children from 4 years up showed a working memory structure similar to the multicomponent model. In structural equation modeling, groups of correlated variables with a common purpose are taken as alternative measures of a particular concept, and models with different plausible causal relations between the represented concepts (called latent variables) are compared to see which model accounts

for the most variability in the data. Other structural equation work leads to the conclusion that the embedded-processes model's focus of attention needs to be considered as one latent variable to capture individual differences in performance on a wider variety of tasks (Gray et al., 2017).

### **The Purpose-of-Model Continuum**

The bottom panel of Figure 2 shows a continuum of some working memory models on the basis of their purposes of study. To the left are models that have taken most of their direct support from idiographic information and have had as a purpose the prediction of individual differences, such as M. Just and Carpenter's (1992) model and earlier supportive work by Daneman and Carpenter (1980; see also Daneman & Merikle, 1996). Engle's (2002) goal maintenance approach is similar except that it has more often included and relied upon a variety of new experimental procedures producing nomothetic results in support of the theory, along with individual differences. Case et al. (1982) and Gaillard, Barrouillet, Jarrold, and Camos (2011) are examples of using developmental data as extreme individual differences, but developmental groups can be considered an intermediate case inasmuch as researchers comparing these groups do not always make as detailed use of individual differences within an age group. The embedded-processes model falls in the middle of the road, depending sometimes on nomothetic results (e.g., Cowan, Saults, & Blume, 2014), other times on developmental differences (e.g., Cowan, Li, Glass, & Saults, 2017; Cowan, Ricker, Clark, Hinrichs, & Glass, 2015), and yet other times on development along with idiographic results within an age group (e.g., Cowan et al., 2005; Cowan, Fristoe, Elliott, Brunner, & Saults, 2006). In the multi-component approach, most of the work has been from the point of view of nomothetic inference, though not wholly without input from idiographic differences, and especially those from cases of brain damage affecting one part of the working memory system or another. Last, on the nomothetic end of the continuum are pioneers, such as James (1890) and Miller (1956), who wrote when it was not yet possible to consider individual differences as precisely as we can do with modern methods.

### **Fruits of Recent Research: Convergence of the Models?**

Why do theorists disagree? There is some disagreement on actual data, but the difference in theories probably comes more from theorists' attention to one or another aspect of a vast literature; it is difficult to consider all of the research at the same time. If we are doing our science well, though, the models should eventually start to converge on the truth. We are happy to report that we think this convergence is happening; changing models are moving toward one another. A key example is that some versions of the multicomponent model and embedded-processes model are becoming more similar, in both a reconciliation between modularity and

attention and a reconciliation between nomothetic and ideographic purposes.

### **Recent Research Reconciling Modularity and Attention**

Although we have presented modularity and attention as separate dimensions of working memory models, there is an intersection between them in that modules supposedly preserve materials with different codes (e.g., verbal and visual-spatial codes) separately, without interference between the two. Attention supposedly allows storage of information from a variety of codes, albeit with the potential for interference between materials from different codes and the potential to prioritize some information at the expense of other information.

### **A Stand-In for Modularity in the Embedded Processing Approach**

In the embedded-processes approach (e.g., Cowan, 1988), there is not a set of different modules (like separate verbal and visuospatial buffers), but there is a prediction similar to models that do have modules. It is the prediction that items with similar features interfere with each other more than do items with different features. In a modular model, this feature-specific interference occurs because items with similar features are held in the same store. In the embedded-processes approach, items, regardless of type, are held in the activated portion of long-term memory, but when items with similar features are concurrently held, they interfere with one another because they depend on the same neural apparatus for that kind of feature. The question for this approach has been how much information is held in the focus of attention versus in the activated portion of long-term memory.

Cowan (1988, 2001) and Saults and Cowan (2007) tended toward the assumption that most information was held in the focus of attention when rehearsal, chunking, grouping processes could not play a role. Further research, however, has led to the changed assumption that, although several items at a time are at first represented in the focus of attention together, they can be quickly off-loaded to the activated portion of long-term memory to free up the attention for other work. Specifically, Cowan et al. (2014) carried out a number of experiments in which a series of verbal items (spoken or printed) were presented along with a spatial array of visual objects. The sets were presented one after another in either order, and participants were required to repeat a single word (*the*) to prevent verbal rehearsal. In some blocks of trials, the task was to remember both sets, and there was a recognition item coming from one set or the other. In other blocks of trials, the task was to remember just one set (verbal in some blocks, nonverbal in others). Using data from these trial types, it was possible to estimate that about two verbal and two visual items could be retained regardless of whether one or both modalities had to be remembered. On top of that, approximately another one item could be retained, with that central capacity devoted

to one modality or split between modalities, depending on the trial type. The explanation was that the focus of attention does not continually retain more than a single item at a time; it may take in and then off-load one set in order to be ready for the second set. The approximately one-item, shared capacity limit may occur for a variety of reasons, such as the need to attend periodically to sets of information in the activated portion of long-term memory in order to refresh or improve the representations. Any such function that would have to be divided between two sets when both of them have to be retained, so refreshing one set comes at the expense of refreshing the other.

### **The Focus of Attention in the Multicomponent Approach**

In the multicomponent modeling approach to working memory, the main role of attention has traditionally been to operate through the central executive to help control cognition. In the current model of A. Baddeley (2000), another possible role is to preserve information via the episodic buffer, which might serve the same role as the focus of attention in Cowan's model (see A. Baddeley, 2001). It is therefore perhaps not surprising that, in recent years, Baddeley, Hitch, and colleagues have investigated the focus of attention and, in particular, priority given to some items in a list at the expense of other list items (Allen, Baddeley, & Hitch, 2017; Hu, Allen, Baddeley, & Hitch, 2016). In these studies, the number of points awarded for recall is set to be greater for some items than for others. There is automatic priority to the last list item, and in addition, participants appear able to prioritize at least one other list item at the expense of other items. Prioritization cannot be simply a matter of encoding of the information, inasmuch as priorities can be set even after the memoranda have disappeared from the computer screen (e.g., Cowan & Morey, 2007; Griffin & Nobre, 2003).

### **Modularity, Attention, and Brain Research**

The interplay between the concepts of attention as a storage device versus nonattentional, possibly specialized storage modes in working memory is a popular theme in recent neuroscientific research on working memory (Cowan et al., 2011; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012; Li, Christ, & Cowan, 2014; Majerus et al., 2016; Reinhart & Woodman, 2014; Rose et al., 2016; Wallis, Stokes, Cousijn, Woolrich, & Nobre, 2015). These neuroimaging studies point to an area in parietal cortex, the intraparietal sulcus, as particularly important in indexing the items held with the help of the focus of attention, whereas the actual neural representation of information in working memory is seen not there but in posterior cortical areas close or identical to the areas in which the initial processing of the information took place. These posterior areas appear to reflect the activated portion of long-term memory or, by another view, modular memory stores along with a partially based focus of attention, whereas central executive control processes appear to rely more heavily on frontal areas. There have thus been leaps in the quest to understand

the neural underpinnings of attention-based and attention-free aspects of working memory.

### **Summary: Reconciling Modularity and Attention**

Across theorists from the multicomponent and embedded-processes camps, there is increasing convergence of their ideas. The embedded-processes camp acknowledges limitations in how much attention is used directly to store information, whereas the multicomponent camp now acknowledges a role of the focus of attention. Still, there are theorists who advocate full modularity (Logie, 2016) or the full use of attention (Morey & Bieler, 2013). With recent technological advances, these mechanisms can be explored more deeply on a neural level.

### ***Recent Research Reconciling Nomothetic and Ideographic Approaches***

During most of the history of working memory research, nomothetic and ideographic approaches have relied on somewhat separate methods. Nomothetic researchers have emphasized careful task analyses, as in most of the research reported by A. D. Baddeley and Hitch (1974). In contrast, ideographic researchers have needed to rely on somewhat standardized tests to examine individual differences in abilities, as in the applications of working memory tests to an understanding of good and poor readers by Daneman and Carpenter (1980; cf. Daneman & Merikle, 1996). In contrast, in recent work on individual differences, careful analyses of certain tasks have proved to be critical for an understanding of individual differences.

Consider, for example, the structural equation modeling work of Gray et al. (2017), fit to 9-year-old children to account for individual differences in performance on a battery of working memory tasks (verbal, spatial, and visual tasks with standard and running span methods). To understand the results, a key task analysis on the basis of past nomothetic work was an analysis of performance on a running digit span task. In each running span trial, participants received a long list of spoken digits without knowing when the list would end. The task was to wait for the end of the list and then recall a small number of the items from the end of the list. There is evidence Gray et al. reviewed that this task is difficult because rehearsal and grouping are not possible (given the long list length and unpredictability regarding when the list will end), making the use of attention critically important for this task. According to other studies that Gray et al. reviewed, nonverbal visual materials also critically require attention for maintenance, whereas rehearsable and groupable verbal materials require less attention. Gray et al. found that list memory tasks with verbal materials were intercorrelated well except for the running verbal span task, running digit span, which, instead, was best intercorrelated with the visual and spatial tasks. This anomaly was resolved using a model in which one latent variable was the focus of attention, subsuming running span along with the visual and spatial tasks. To provide the best fit, the multicomponent model had to be modified to be more like the

embedded-processes model, replacing the visual-spatial store with storage in the focus of attention. Thus, the task analysis from previous nomothetic work contributed to an understanding of individual differences in working memory.

The nomothetic analysis of many tasks by Unsworth and Engle (2007) indicated that there is no fundamental difference between short-term memory tasks that required only storage and working memory tasks that required both storage and processing. It was aspects of both kinds of tasks requiring the control of attention that distinguished between better and poorer performers, no matter whether attention was needed for maintenance, manipulation, or long-term retrieval. As a final example of nomothetic analysis contributing to ideographic knowledge, Friedman et al. (2006) used an analysis of executive function into three more specific functions and found that updating of information in working memory correlated with intelligence; shifting of the focus of attention and inhibition of irrelevant material did not.

The examination of different groups is an important part of the ideographic approach, and it, too, benefits from careful task analyses in recent work. Cowan (2016, 2017b) summarized work addressing the question of what accounts for developmental differences in working memory. Task analyses were used to equate participants in age groups from the early elementary school years to adulthood on various confounding processes, such as the ability to ignore distracting items, the ability to rehearse, and familiarity with the information. In every case, working memory still increased with age even with these factors equated. Cowan et al. (2017), however, learned more by carrying out a study comparable to the adult work by Cowan et al. (2014), requiring working memory maintenance of both arrays of colored spots and series of tones. They found that what improved with development was not the ability to hold more information in the focus of attention but the ability to preserve more information from each modality in a manner making it more resistant to cross-modality interference. Perhaps, it is the ability to off-load information efficiently to the activated portion of long-term memory that improves with development. In another example, coming from the view that what distinguishes younger versus more mature participants is the speed at which items can be mentally refreshed in working memory. Gaillard et al. (2011) carefully manipulated materials to equate that speed across age groups. In that way, they were able to equate working memory performance levels across age groups. Combining these studies, we do not yet know whether working memory increases with age because processing speed increases allow faster refreshing of items before they decay or, conversely, whether processing speed increases because a larger working memory capacity allows more items to be refreshed in parallel (Lemaire, Pageot, Plancher, & Portrat, 2017). Moreover, what is called *refreshing* might actually be the successful off-loading of information out of the focus of attention and into activated long-term memory. It could entail the use of attention to discover and memorize structure, forming groups and chunks and, thereby, reducing the rate of forgetting.

### **Summary: Reconciling Nomothetic and Ideographic Approaches**

A very exciting development in the field of working memory is the coming together of careful task analysis and nomothetic information and the examination of individual and group differences, all in the same studies. In 2017, Randall Engle gave the keynote address for the annual Psychonomic Society meeting essentially on the basis of the premise that there has been a long-standing need to merge the nomothetic and ideographic approaches, which he illustrated in the field of working memory.

### **Conclusion: Potential New Directions for Language Research**

The theories of working memory are leading us closer to an understanding of the extent to which language, like other information, is retained in separate modules versus a common problem-solving space, how much it depends on attention as opposed to automatic processing, and how much it can benefit from ideographic and nomothetic experimentation. In future work, we might anticipate that a new understanding of these themes in the field of working memory can be applied to more connected discourse. For example, we might learn more about the role of attention and working memory in the misinterpretation of sentences (e.g., Patson et al., 2009) and learn who is most susceptible to these misinterpretations and under what conditions. We might learn whether the degree of modularity (or non-modularity) is similar for working memory and language. Finally, we might learn what language mechanisms change with normal and abnormal development and how individual differences in language may depend on working memory capabilities. In short, the field is thriving.

### **Acknowledgments**

This work was completed with support from National Institutes of Health Grant R01 HD-21338 to Cowan.

### **References**

- Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology, 97*, 79–97.
- Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2017). Executive and perceptual distraction in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance, 43*(9), 1677–1693.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory, 2* (pp. 89–195). New York, NY: Academic Press.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*, 622–628.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

- Baddeley, A.** (2001). The magic number and the episodic buffer. *Behavioral and Brain Sciences*, *24*, 117–118.
- Baddeley, A.** (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*, 189–208.
- Baddeley, A. D.** (1986). *Working memory*. Oxford, United Kingdom: Clarendon Press.
- Baddeley, A. D., & Hitch, G.** (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C.**, (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.
- Baddeley, A. D., Thomson, N., & Buchanan, M.** (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.
- Barrouillet, P., Bernardin, S., & Camos, V.** (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.
- Bjork, R. A., & Whitten, W. B.** (1974). Recency—Sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*, 173–189.
- Broadbent, D. E.** (1958). *Perception and communication*. New York, NY: Pergamon Press.
- Camos, V., Mora, G., & Oberauer, K.** (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition*, *39*, 231–244.
- Carpenter, P. A., Miyake, A., & Just, M. A.** (1994). *Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging*. San Diego, CA: Academic Press.
- Case, R., Kurland, D. M., & Goldberg, J.** (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386–404.
- Conrad, R.** (1964). Acoustic confusion in immediate memory. *British Journal of Psychology*, *55*, 75–84.
- Cowan, N.** (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human maturation processing system. *Psychological Bulletin*, *104*, 163–191.
- Cowan, N.** (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, *31*, 668–684.
- Cowan, N.** (1995). *Attention and memory: An integrated framework*. *Oxford Psychology Series, No. 26*. New York, NY: Oxford University Press.
- Cowan, N.** (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge, United Kingdom: Cambridge University Press.
- Cowan, N.** (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
- Cowan, N.** (2000/2001). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting*, *5*(2), 117–146.
- Cowan, N.** (2015). Second-language use, theories of working memory, and the Vennian mind. In Z. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 29–40). Bristol, United Kingdom: Multilingual Matters.
- Cowan, N.** (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science*, *11*, 239–264.
- Cowan, N.** (2017a). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170.
- Cowan, N.** (2017b). Mental objects in working memory: Development of basic capacity or of cognitive completion? *Advances in Child Development and Behavior*, *52*, 81–104.
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A.** (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*, 42–100.
- Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S.** (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, *34*, 1754–1768.
- Cowan, N., Hogan, T. P., Alt, M., Green, S., Cabbage, K. L., Brinkley, S., & Gray, S.** (2017). Short-term memory in childhood dyslexia: Deficient serial order in multiple modalities. *Dyslexia*, *23*(3), 209–233.
- Cowan, N., Li, D., Moffitt, A., Becker, T. M., Martin, E. A., Saults, J. S., & Christ, S. E.** (2011). A neural region of abstract working memory. *Journal of Cognitive Neuroscience*, *23*, 2852–2863.
- Cowan, N., Li, Y., Glass, B., & Saults, J. S.** (2017). Development of the ability to combine visual and acoustic information in working memory. *Developmental Science*. Advance online publication. <https://doi.org/10.1111/desc.12635>
- Cowan, N., & Morey, C. C.** (2007). How can dual-task working memory retention limits be investigated? *Psychological Science*, *18*, 686–688.
- Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A.** (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science*, *18*, 132–145.
- Cowan, N., Saults, J. S., & Blume, C. L.** (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General*, *143*, 1806–1836.
- Daneman, M., & Carpenter, P. A.** (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450–466.
- Daneman, M., & Merikle, P. M.** (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*(4), 422–433.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarman, H. J., & Usher, M.** (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3–42.
- de Jong, P. F.** (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology*, *70*, 75–96.
- Engle, R. W.** (2002). Working memory capacity as executive attention. *Current directions in psychological science*, *11*(1), 19–23.
- Ferreira, F., & Clifton, C.** (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K.** (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172–179.
- Gaillard, V., Barrouillet, P., Jarrold, C., & Camos, V.** (2011). Developmental differences in working memory: Where do they come from? *Journal of Experimental Child Psychology*, *110*, 469–479.
- Gathercole, S. E., & Alloway, T. P.** (2006). Practitioner review: Short-term and working memory impairments in neurodevelopmental disorders: Diagnosis and remedial support. *The Journal of Child Psychology and Psychiatry*, *47*(1), 4–15.
- Gathercole, S. E., & Baddeley, A. D.** (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*, 336–360.

- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177–190.
- Gillam, R. B., Cowan, N., & Day, L. S. (1995). Sequential memory in children with and without language impairment. *Journal of Speech and Hearing Research, 38*, 393–402.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior, 5*, 351–360.
- Gray, S., Green, S., Alt, M., Hogan, T., Kuo, T., Brinkley, S., & Cowan, N. (2017). The structure of working memory in young children and its relation to intelligence. *Journal of Memory and Language, 92*, 183–201.
- Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience, 15*, 1176–1194.
- Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11*, 236–242.
- Hu, Y., Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2016). Executive control of stimulus-driven and goal-directed attention in visual working memory. *Attention, Perception, & Psychophysics, 78*, 2164–2175.
- Jalbert, A., Neath, I., & Surprenant, A. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition, 39*, 1198–1210.
- James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228–238.
- Just, M., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149.
- Kane, M. J., & Engle, R. W. (2000). Working memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 336–358.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. E. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.
- Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*, 153–161.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580–602.
- Lemaire, B., Pageot, A., Plancher, G., & Portrat, S. (2017). What is the time course of working memory attentional refreshing? *Psychonomic Bulletin & Review, 25*(1), 370–385. <https://doi.org/10.3758/s13423-017-1282-z>
- Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience, 24*, 61–79.
- Li, D., Christ, S. E., & Cowan, N. (2014). Domain-general and domain-specific functional networks in working memory. *NeuroImage, 102*, 646–656.
- Logie, R. H. (2016). Retiring the central executive. *The Quarterly Journal of Experimental Psychology, 69*(10), 2093–2109.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17*, 347–356.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35–54.
- Majerus, S., & Cowan, N. (2016). The nature of verbal short-term impairment in dyslexia: The importance of serial order. *Frontiers in Psychology, 7*, 1522. <https://doi.org/10.3389/fpsyg.2016.01522>
- Majerus, S., Cowan, N., Peters, F., Van Calster, L., Phillips, C., & Schrouff, J. (2016). Cross-modal decoding of neural patterns associated with working memory: Evidence for attention-based accounts of working memory. *Cerebral Cortex, 26*, 166–179.
- Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 46*, 1138–1153.
- McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39*(4), 352–370.
- Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1–21.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York, NY: Holt, Rinehart & Winston.
- Milner, B. (1968). Disorders of memory after brain lesions in man. *Neuropsychologia, 6*, 175–179.
- Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of Communication Disorders, 36*, 221–231.
- Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires attention. *Psychonomic Bulletin & Review, 20*, 163–170.
- Murdock, B. B. (1961). The retention of individual items. *Journal of Experimental Psychology, 62*, 618–625.
- Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*, 53–81.
- Neath, I., & Brown, G. D. A. (2012). Arguments against memory trace decay: A SIMPLE account of Baddeley and Scott. *Frontiers in Psychology, 3*, 35.
- Newell, A., & Simon, H. A. (1956). *The logic theory machine: A complex information processing system*. Santa Monica, CA: Rand Corporation.
- Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115*, 544–576.
- Oberauer, K., & Lin, H. (2017). An interference model of visual working memory. *Psychological Review, 124*(1), 21–59.
- Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 280–285.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.
- Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex, 43*, 135–145.
- Reinhart, R. M., & Woodman, G. F. (2014). High stakes trigger the use of multiple memories to enhance the control of attention. *Cerebral Cortex, 24*, 2022–2035.

- Ricker, T. J.** (2015). The role of short-term consolidation in memory persistence. *AIMS Neuroscience*, 2(4), 259–279. <https://doi.org/10.3934/Neuroscience.2015.4.259>
- Ricker, T. J., & Cowan, N.** (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 417–428.
- Ricker, T. J., & Hardman, K. O.** (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General*, 146(11), 1551–1573. <https://doi.org/10.1037/xge0000346>
- Ricker, T. J., Spiegel, L. R., & Cowan, N.** (2014). Time-based loss in visual short-term memory is from trace decay, not temporal distinctiveness. *Journal of Experimental Psychology*, 40(6), 1510–1523.
- Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R.** (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, 354, 1136–1139.
- Saults, J. S., & Cowan, N.** (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, 136, 663–684.
- Schneider, W., & Detweiler, M.** (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 57–70). New York, NY: Academic Press.
- Schweickert, R., Guentert, L., & Hersberger, L.** (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, 1, 74–77.
- Simons, D. J.** (2000). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, 4, 147–155.
- Swanson, H. L.** (1999). What develops in working memory? A life-span perspective. *Developmental Psychology*, 35, 986–1000.
- Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F.** (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136, 64–81.
- Thorndike, E. L.** (1914). *The psychology of learning*. New York, NY: Teachers College.
- Underwood, B. J.** (1957). Interference and forgetting. *Psychological Review*, 64, 49–60.
- Unsworth, N., & Engle, R. W.** (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104–132.
- Vandierendonck, A.** (2016). A working memory system with distributed executive control. *Perspectives on Psychological Science*, 11, 74–100.
- Vergauwe, E., Barrouillet, P., & Camos, V.** (2010). Do mental processes share a domain general resource? *Psychological Science*, 21, 384–390.
- Wallis, G., Stokes, M., Cousijn, H., Woolrich, M., & Nobre, A. C.** (2015). Frontoparietal and cingulo-opercular networks play dissociable roles in control of working memory. *Journal of Cognitive Neuroscience*, 27, 2019–2034.
- Waters, G. S., & Caplan, D.** (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter. *Psychological Review*, 103(4), 761–772.
- Weismer, S. E., Evans, J., & Hesketh, L.** (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 42, 1249–1260.