Advanced Review

Working memory

Timothy J. Ricker, Angela M. AuBuchon and Nelson Cowan*

The working memory system maintains the limited information that can be kept in mind at one time. These memories are distinct from the vast amount of information stored in long-term memory. Here we give a brief summary of findings over the past half-century in the areas of working memory that we see as particularly important for understanding its nature. We discuss several current controversies, including whether there are different systems or brain modules for different kinds of working memory, why we lose items from working memory, and how individuals and age groups differ. We try to describe what is and is not known. Last, a discussion of findings from neuroimaging helps to constrain working memory theory. © 2010 John Wiley & Sons, Ltd. *WIREs Cogn Sci*

Discussions of working memory often suffer from the wide variety of usages of the term. Some have described working memory literally, as information maintained temporarily and used to accomplish mental work.[1] Others have used the term more specifically in reference to a proposed multi-component system that temporarily maintains information and manipulates it to carry out cognitive tasks.[2] In the present article, when we discuss working memory, we are referring to both the information that can be kept available in mind at any given time and the processing that occurs to maintain this information, but not other processing.

The amount of information that can be maintained for quick and easy access is clearly limited. This is apparent when you try to remember directions that have too many turns. Imagine that you are leaving work for the day and realize that you do not have directions to this evening's celebration at an acquaintance's house. Because you are on the way out and you do not have any paper to write down the directions when a coworker tells them to you. Instead, you must remember them until you drive home and can write them down: left on Fifth Street, right onto Elm, left onto Bircham, left onto Lupine, right onto Ridge, third house on the right. When you get home and write the directions down you would probably find that you have a hard time remembering a few parts, like the name of the third- or fourth-mentioned street, or whether you turned right or left onto elm. You may even forget the whole thing if you get in a conversation during the drive. To succeed you need to write down the directions, and even after doing so working memory is needed to execute the next turn in as much, as one cannot stare at the directions constantly while driving. As another example, while doing arithmetic mentally, one must hold partial results in mind while doing calculations. A little reflection on these two examples tells us that the number of turns, digits, or other items that can be held in conscious memory is quite small. The items held in this limited cognitive space are said to be in working memory.

Working memory is important because it mediates most of our conscious interactions with the world. On a more practical level, working memory is critical for a wide variety of cognitive functions. Because it holds information that is being processed in an available state, its size and functioning affect how we are able to think about and solve problems. Individual differences in working memory are predictive of performance on complex cognitive tasks, such as reasoning,[3] and clinical deficits in working memory are related to conditions, such as attention deficit disorder[4] and schizophrenia.[5]

It is agreed virtually by all cognitive psychologists that the processes attributed to working memory are essential in human cognition. One must keep information in mind while processing it to function intellectually and socially. There are, however, still fundamental issues and disagreements regarding just how working memory is limited, and how it operates. We have chosen to present the research in this area with the controversies firmly in mind and therefore organize the discussion around these controversies. The questions we consider include, whether working

*Correspondence to: CowanN@missouri.edu

University of Missouri, 210 McAlester Hall, Columbia, Missouri, USA

DOI: 10.1002/wcs.50

memory is a single system or is composed of several independent systems, why there is a limit to the number of items that can be held in memory, whether or not the passage of time causes us to forget, why the capacity of working memory differs among individuals, and why it changes with development. We concentrate on behavioral research but conclude with a glimmer of the recent brain imaging research and how it relates to more traditional behavioral research.

IS WORKING MEMORY ONE SYSTEM OR SEVERAL SYSTEMS?

Although researchers at first conceived of working memory as a single system or process,[6,7] many have subsequently thought that the data compel us to propose that there are multiple systems or processes that are involved in our ability to retain a small amount of information in a highly accessible state.[8–11] In this section we consider whether there are separate visual and verbal modules for working memory; a general, central faculty; or both general and modality- or code-specific aspects of working memory. To preview our conclusion, we believe that there is a general working memory and that some aspects of working memory also are code-specific, although these code-specific properties do not necessarily constitute separate systems. Here we briefly summarize the history of this key issue and present evidence that the central memory is limited in terms of the number of chunks. Later, we will examine other kinds of working memory limits for code-specific information, in particular time-based forgetting.

Miller[6] famously reviewed evidence that the amount that can be remembered and repeated is limited to about seven items, give or take a few depending on the type of items that are to be remembered and the particular person. In the early days of cognitive psychology, this phenomenon was illustrated within models of human information processing by a single box that represented the limited number of items that could be recalled (e.g., Ref 7; Figure 1(a)). Miller also discussed a trick that individuals could use to increase their memory span: items could be grouped together to form larger chunks that act like individual items. For example, one can recall nine random letters easily if they form acronyms, such as IBM, CIA, FBI. Another trick that was discussed later was silently repeating items to one's self or rehearsing them.[8]

Cowan[9] reviewed research indicating that, when one prevents chunking and rehearsal, the number of independent items that can be recalled is typically only about 3–5. It remains unclear, however, why

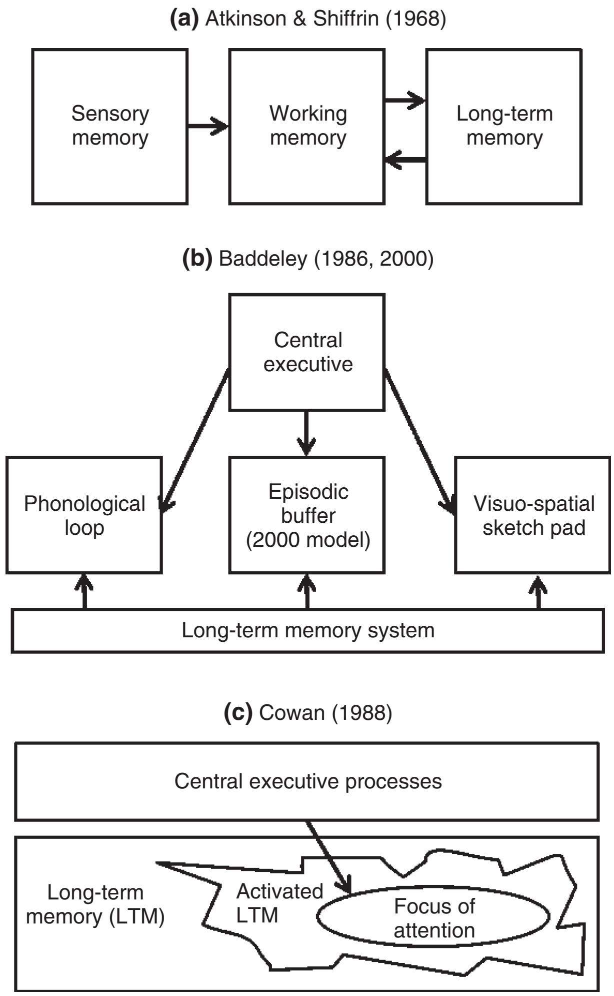

FIGURE 1 | Models of working memory as conceptualized by three researchers: (a) Atkinson and Shiffrin.[7] (b) Baddeley.[8,10] (c) Cowan.[11]

such strict capacity limits occur. To try to find out, we will need to investigate further the nature of the working memory system or systems, as the case may be.

In a seminal article that has guided the field for many years, Baddeley and Hitch[2] reported on many experiments suggesting that there is less interference between tasks than one would expect on the basis of the single-system working memory box model of Atkinson and Shiffrin.[7] For example, Baddeley and Hitch found that remembering a series of digits interfered with verbal reasoning or comprehension, but only if the number of digits was challenging (six digits, but not three). They also demonstrated that verbal distraction interfered with verbal memory more than did visual distraction, and that visual distraction interfered with visual memory more than did verbal distraction. On the basis of this information, Baddeley and Hitch theorized that working memory has a central storage resource where information could be held while it was needed for processes such as reasoning and comprehension. However, they also thought that there were other types of storage (verbal-phonological and visual-spatial) that were not shared with processing and were separate from one another. Small verbal memory loads did not interfere with reasoning or comprehension, the explanation went, because the information could be saved as a series of speech sounds in a phonological store that is separate from what is used for reasoning or comprehension. Baddeley[8] crystallized the model so that it included a phonological loop (so called because it involved covert rehearsal of the phonological information in a repeating loop) and a visuo-spatial sketchpad operating similarly. He also removed the central store for the sake of parsimony, though he later added one back, called an episodic buffer[10] (Figure 1(b)). This left a central executive carrying out all the processing and managing the storage of all the information in the phonological and visuo-spatial stores. Whereas the central executive was attention-demanding, in this model the stores themselves were attention-free, including the rehearsal process. It is not yet clear if attention is needed for the proposed episodic buffer.

Cowan[11] was uncomfortable with the division of labor in the Baddeley[8] model. Although he accepted that there is more interference between like kinds of items in working memory, it did not seem likely that the only important distinctions are between verbal-phonological and visual-spatial information. How could we understand, for example, memory for tones in various spatial locations or memory for touch information? In the absence of definitive knowledge of the taxonomy of stores, Cowan lumped together many kinds of information storage and hypothesized that they are all instances in which information from the vast banks of long-term memory are temporarily in an activated state. Also in contrast to the Baddeley[8] model, Cowan proposed that a limited amount of information is also held in the focus of attention (Figure 1(c)) rather than simply held in passive stores. Whereas the activated memory was supposed to be limited by the amount of time that has passed and the types of interference that have occurred, the information in the focus of attention was supposed to be limited to just a few (3–5) separate items, a theme that Cowan[9] developed further.

All of the models after the era of Atkinson and Shiffrin[7] seem to postulate more than one working memory storage system (verbal-phonological and visual-spatial in Figure 1(b); activated memory and focus of attention in Figure 1(c)). One critical point that distinguishes the models is whether there exists a central, abstract kind of working memory in which information from any source can be entered, i.e., a *domain-general* working memory store. Baddeley[8] said no; Cowan[11] and Baddeley[10] said yes. Cowan further proposed that the domain-general store is the focus of attention and therefore is an especially noteworthy form of working memory.

Cocchini, Logie, Della Sala, MacPherson, and Baddeley[12] conducted a well-designed study in which there were two tasks on some trials, often including a verbal memory task and a spatial pattern memory task. There was little interference between these different tasks (i.e., little cross-modal interference), leading Cocchini et al. to reject the idea of a domain-general store. That was the case in other dual-task studies as well (e.g., Ref 13). Yet, other studies with slightly different designs have shown more substantial cross-modal interference between visual and verbal materials.[14–17] Theorists accept that there is some such interference but differ dramatically in the explanation for that interference.

According to Baddeley,[8] the interference should not occur. According to Baddeley[10] or Cowan[9,11] the interference could occur in a domain-general working memory store. Cowan goes further in predicting that if one could get rid of most sources of activated memory, such as memory for the way the items look or sound, what is left over is a store limited to a fixed number of conceptual items. Saults and Cowan[18] addressed this issue in an experiment in which participants received spoken digits and arrays of visual colored spots. In order to make rehearsal difficult, the spoken digits were presented simultaneously from four different loudspeakers in different voices. Participants were sometimes responsible for only one sensory modality but at other times they were responsible for both. It turned out that a fixed limit occurred provided that sensory memory activation was eliminated with a mask. The mask was a combination of multicolored spots and garbled speech, which interfered with how the memory of the stimuli looked and sounded. After the mask, there was a repetition of the stimuli to be remembered, but sometimes with a change in one item. When only visual memory was required, people could remember 3–4 visual items on average. When auditory memory was also required, people could still remember a total of 3–4 items, but now some of those items were the spoken digits and others were the colored spots. This seems like good evidence for a domain-general memory.

Perhaps the most difficult part of a domain-general memory to prove is that the domain-generality comes in the form of storage as opposed to processing. If attention is needed to carry out a refreshing process (similar to thinking of an item, thereby reactivating its representation) for any stimuli[19,20] then the limit could be in the capability of the refreshing process rather than the contents of the focus of attention. We cannot yet clearly distinguish between these hypotheses.

Other research methods do support the notion that attention is involved in storing information from multiple modalities. Just et al.[14] had participants perform a verbal sentence comprehension task and a visual mental rotation task and observed dual-task deficits in performance relative to when the tasks were performed alone. In this experiment functional magnetic resonance imaging (fMRI) data were also obtained and demonstrated tradeoffs similar to those found in the behavioral data. When performed alone, the verbal and visual tasks activated separate brain areas in the temporal and parietal lobes, respectively, indicating separate resources for memory representations. When both tasks were performed simultaneously, the same areas were again activated, but activation levels were significantly lower in both of these areas in the dual-task situation compared to when the tasks were performed individually. This suggests that a common resource, namely attention, had to be shared between the modality-specific representations in the temporal and parietal cortex, consistent with the theory that a common attention-related resource influences both modalities.

In a study examining event-related brain potentials during a dual-task procedure, the amplitude of the P300 component elicited by an auditory secondary task was shown to decrease as the difficulty of a visual primary task increased.[16] This is important because the P300 component is considered to be a measure of working memory updating, and is not affected by response selection or execution.[21] The decrease in P300 amplitude for the secondary task suggests that increasing the resources required to perform a visual task decreases the amount of resources available to perform a concurrent auditory task. Critically, Sirevaag et al.[16] also found that the visual task P300 increased as the visual task increased in difficulty and that the magnitude of the decrease in the auditory secondary task P300 was complementary to the increase in magnitude for the visual primary task, consistent with the use of a domain general resource for task performance.

Statistical analysis of individual differences has also been used to support theories of a domain general working-memory component. Kane et al.[22] performed factor analysis of 12 tasks, 3 from each of the following categories, verbal working memory, visual working memory, verbal short-term memory, and visual short-term memory and examined how well they predicted general fluid intelligence (gF), the ability to solve new problems. The best latent variable model explaining variability in these tasks was composed of a general working memory capacity with contributions from spatial and verbal long-term memory. Working memory capacity then predicted gF. This organization describes a domain-general working memory model in which information from all modalities is utilized by common maintenance and processing components while retaining its modality specific attributes.

If we are correct that a domain-general capacity-limited working memory store exists it does not mean that working memory cannot also contain information that is specific to a modality (e.g., how a word looks vs. how it sounds). Also, certain types of special processing may be specific to some kinds of stimuli; covert verbal rehearsal, for example, can help in the recall of the serial order of items only if they can be labeled easily. It would be easy to rehearse a phone number but not the shape of an irregular polygon. These possible modality-specific features of working memory do not negate the idea that there is a domain-general form of storage. The discussion of whether this domain-general resource exists, or if working memory is instead several independent, yet related, systems is clearly critical to understanding the organization of the brain and our own limits. Encouragingly, it appears that progress is being made towards a consensus on some sort of general-resource for all modalities. The fact that we have a limited central working memory system, however, naturally leads to the question, why is this system limited?

WHY CAN WE RECALL ONLY A LIMITED NUMBER OF CHUNKS OF INFORMATION?

Earlier, we stated that working memory capacity is generally limited to 3–5 meaningful items or chunks. A good example of this constant capacity limit is demonstrated by Chen and Cowan.[23] They presented lists of single words or learned pairs of words for serial recall. They prevented rehearsal and its contributions to performance through articulatory suppression, a procedure in which a single word or sound is continuously repeated (e.g., 'the, the, the...'). Capacity was measured in chunks, a name for meaningful units of information, which could either be a single word or a pair of words, depending on how items were presented. In line with past estimates,[9,24] Chen and Cowan consistently observed recall of about 3 chunks across lists of 6–12 words presented as either single words or word pairs. This study illustrates that, when rehearsal and chunking of items are accounted for, the contents of working memory are observed to be a fairly constant number of conceptual items.

Although it is clear that this capacity limit exists, why it exists is an open question without a single answer. Explanations of why we have a capacity limit include functional and mechanistic answers. In the present section, we give brief descriptions of several of these explanations that we find most likely. Although each view will provide a different answer to why we have a fundamental capacity limit, they are generally complementary rather than competing explanations.

The functional explanations treat the capacity limit as a strength. We cannot process everything in the world perfectly so there has to be a decision about what to process, and attention is given to some information at the expense of other information. The capacity limit in working memory may make that decision possible by keeping a limited few items protected from proactive interference, or confusability with previous items, at any given time.[25,26] Similarly, there is a limit in the number of links or bindings that can be formed between one item and another or between an item and its context.[27] The process of evolution[28] could have favored individuals with a working memory size that is well-suited to the tasks that individuals must carry out. Smaller working memories would be insufficient for self-preservation and larger working memories would be biologically too costly to maintain.

In another sort of functional argument, it has been suggested that the size of working memory we have is not only the best that can be done with the amount of biological energy available; but that it is, in fact, optimal. Dirlam[29] and MacGregor[30] put forward mathematical arguments based on certain assumptions about how people search through memory. The search process involves finding the right group or chunk of information and then finding the right item within the group. With this kind of analysis it turns out that the size of grouping that is most efficient is within the range of the number of items in working memory. So assuming that items that are in working memory at the same time can be grouped together, our working memory could have evolved to allow maximally efficient groups to be formed and later searched.

The mechanistic explanations address the issue of just what it is in the brain that allows a certain number of items to be remembered at once and prevents more from being remembered. One possibility is based on Milner's[31] proposition that memory items exist because the neurons that represent their features fire synchronously. This idea has merit, as several studies have found that synchronous firing at around 40 Hz seems to define an object[32] and that this pattern of neuronal function seems to be related to attention.[33] Lisman and Idiart[34] reviewed neural evidence suggesting that the synchronous firing for all items in working memory has to be repeated within a period of about a tenth of a second. If too many items are represented within that period then the features of one item can be confused with the features of another; a red circle and blue square can be misremembered as a red square and blue circle, for example. Other recent work has shown that areas of the parietal cortex demonstrate activation that reflects the number of items that can be kept in mind at any given time.[35,36] Perhaps the synchronous firing occurs in relation to these areas, consistent with some previous findings.[33] More research is certainly required to validate these theories, but they provide interesting leads toward determining why we are constrained in much of our thinking.

Whereas we thus believe that the central component of working memory is limited in the number of chunks it can hold at once, we and others have proposed that modality- and code-specific aspects of working memory are limited in the duration for which activation of memory can persist.[8,11] Perhaps time limits also apply to central memory. The next section explores the possibility of time limits in working memory.

WHY DOES WORKING MEMORY APPEAR TO HAVE A TIME LIMIT?

As we have discussed, our working memory is limited in capacity, and doing, thinking, and remembering all require some of this limited capacity in order to function. If we do not pay special attention to the items we wish to remember, such as the digits in a phone number we have just been told, they seem to evaporate from our minds very quickly. Peterson and Peterson[37] showed that memory for consonant trigrams (such as M-K-V) became worse over an 18 s time period while participants counted backwards by 3 or 4 s from a random three-digit number, eventually resulting in almost no remaining memory of the trigram. This demonstrates just how quickly we forget even a small amount of information held in mind while distracted by another task. In the present section, we discuss the two prominent explanations for why we forget over time in situations like this one.

One description of why we forget claims that items are represented in the brain as activated memory traces and that this activation decays with time if the items are not mentally rehearsed or refreshed.[8,11,20,38,39] At some point, the activation level becomes too low and items are no longer retrievable for conscious recall. This decay theory would explain the Peterson and Peterson results as at least partially due to the passage of time. The opposing position is that once items are in memory they remain until some process interferes with the item representation through displacement, feature overwriting, or some other process,[40,41] at which point information is lost and forgetting occurs. Interference theories would explain the Peterson and Peterson findings as entirely the result of interference from the counting task and buildup of item familiarity over multiple trials.

Traditionally, the strongest evidence supporting temporal decay of memory traces as the reason for forgetting has been the word length effect (WLE), a term meaning that fewer longer words can be remembered concurrently than shorter words.[42] This was supposedly because memory traces that are not rehearsed will decay beyond the threshold for recovery within about 2 s, and longer words require more time per word to rehearse. Although the WLE has been replicated many times,[43–45] a number of confounds eventually discredited it as firm evidence for time-based decay. Especially damaging was evidence that increased phonologic complexity, not articulatory duration, may underlie the WLE.[46] Currently there are arguments both supporting the WLE as a result of memory decay[45] and denying that the WLE provides evidence for temporal decay,[47] leaving its status uncertain.

More direct tests of decay have also been tried. Cowan and AuBuchon[48] presented lists with an irregular timing and sometimes required that recall also be in that timing. The theoretical rationale was that it is difficult to rehearse while remembering the timing. It was found that lists with long intervals near the beginning impaired recall of items later in the list, in keeping with a decay hypothesis. Other studies, however, have directly manipulated the pace of recall and have not found very much impact of the recall pace on the amount recalled.[49–51]

Perhaps the strongest current argument for a time-based decay explanation of forgetting comes from research by Barrouillet and colleagues[20,38] using the continuous span task paradigm to support their time-based resource-sharing (TBRS) model of working-memory. The TBRS model consists of several critical assumptions. First, items are activated memory traces which decay quickly and are forgotten. These items are maintained through an attention-based refreshing mechanism which is also necessary for performing long-term memory retrievals. Attention may be capable of performing only one task at a time, which leads to quicker loss of memory items when a more difficult or more frequent secondary task must be carried out. These premises lead to the predication that the proportion of time occupied by a distracting nonmaintenance task, compared to total time that the memory must be maintained, predicts the amount of forgetting. Evidence in favor of the TBRS model comes primarily from experiments in which participants are to remember a series of items, such as words, digits, or letters. After the presentation of each item, a processing task is presented, such as reading numbers off a screen or solving a simple math equation. Once all items to be recalled and related processing events have been presented, the memory items are all to be recalled. The ratio of nonmaintenance time to total time available, called the cognitive load, has a negative linear relationship with the accuracy of recall.[20,38,52] This supports the notion of decay unless another explanation can be found as to why recall is linearly related to cognitive load.

In contrast to memory decay, the existence of forgetting due to various forms of interference effects is not a controversial topic. In fact, most decay theorists readily agree that a large amount of forgetting within working memory is due to interference. Interference theorists, however, differ from decay theorists in that they believe that all forgetting within working memory, not just some or most, is due to interference effects. Oberauer and Kliegl,[41] for example, mathematically model performance specifically as a function of feature overlap between items, speed of processing, and level of nonrelevant concept activation. These parameters are allowed to vary both across participants and across tasks in order to reflect effects of both individual differences and varying task requirements. This conceptualization appears to be fairly successful, explaining more than 80% of observed variance in at least one verbal and one spatial task. In these models, working memory capacity, though a domain-general concept, is not constant. The level of overlap between currently held representations will cause a predictable amount of forgetting, resulting in lower functional capacity for relatively more similar representations. It remains to be seen whether this model can accommodate the results favoring a constant capacity (e.g., Refs 9,18,23).

Oberauer and Kliegl also claim that their model can explain the findings of Barrouillet and colleagues by positing that free time during memory retention is spent repairing damaged memory traces. The critical factor explaining memory performance would be free time for repair, not the time over which memory decays as in the TBRS account. These competing claims were investigated by Portrat, Barrouillet, and Camos.[52] They varied the amount of time between items during which distracting processing had to be carried out but left constant the amount of time without distraction. They found that more time processing a distraction still meant more memory loss, in keeping with the decay theory. Given that a single study rarely can resolve a basic issue, more investigation is needed to determine which view of forgetting is correct, or if some hybrid model should be preferred.

In sum, all investigators clearly agree that information is lost over time, but some investigators attribute this solely to the effects of interference from additional material that is processed during a retention interval, whereas other investigators believe that there are some effects of time regardless of the nature of interfering stimuli.

The previous sections of this review have discussed how everyone's working memories are the same in several fundamental, though debated, ways. The following two sections review findings of differences across people in working memory function and what this may say about the fundamental structure of working memory in general.

HOW AND WHY DO PEOPLE DIFFER IN WORKING MEMORY ABILITIES?

Given that the estimated limits of the working memory capacity range from only 3–5 items, it seems surprising that individual variation would be large enough to influence performance on other tasks. Yet this is exactly what we observe, space for one less item in working memory could mean forgetting one important point in a conversation or one key part of a problem. We start by describing working memory span tasks and their relationship to other cognitive abilities. Then we discuss several proposed explanations of how storage and processing contribute to overall working memory capacity.

The capacity limited, domain-general form of memory we have discussed above appears closely linked to cognitive ability. Daneman and Carpenter[53] asked participants to read sets of sentences, while remembering the last word of each sentence. College students in their study were able to remember only 2–5 words, depending on the individual. This surprisingly low memory ability relative to the estimates of earlier research[6] is understandable because attention is occupied with processing between the presentation of items, preventing the grouping of multiple words into a chunk. This reading span measure of Daneman and Carpenter was strongly correlated with verbal scholastic aptitude test (SAT) scores, as well as several other reading comprehension measures, whereas memory for lists of words was not significantly correlated with the same comprehension measures. Later research also uncovered a strong relationship between working memory tasks requiring both storage and processing and a large variety of other measures of complex cognition, including, for example, learning to spell,[54] reasoning ability,[3] and gF.[55] These effects were obtained even when the working memory task included spatial processing along with verbal storage, suggesting that a domain-general working memory resource is used in these tasks and in complex cognitive functions.[22]

The question remains, what drives these individual differences between people? One possibility is that individuals who demonstrate higher working memory spans have more efficient executive functions, so that the processing task consumes less attention and leaves more for storage.[56] A second theory posits that individual differences in both processing and storage capacity can contribute to overall differences in working memory performance.[57]

Various tasks show the role of processing ability. For example, Conway, Cowan, and Bunting[58] used a dichotic listening task intended to replicate the 'cocktail party phenomenon' in which people tend to hear their own name amid a noisy environment. They found that only 20% of high span individuals (in the upper quartile of working memory performance) noticed that their name had been presented in the unattended auditory channel, compared to 65% of low span individuals. This difference is likely due to the high spans' better ability to block the irrelevant information and selectively attend to a single auditory channel. High span individuals' ability to better selectively attend to, or selectively inhibit, information has also been demonstrated in the visual modality. Vogel, McCollough, and Machizawa[59] showed this with memory for visual arrays using an event-related potential measure of working memory load, the contralateral delay activity (CDA) amplitude. Participants saw a visual array containing target and distracter items. In one task, for example, they were to remember the orientations of green bars and ignore the orientations of red bars. Low-spans' CDA amplitude indicated that they were maintaining both the target and distracter items in working memory, but the lower CDA amplitude of high-span individuals indicated that they were maintaining only the target items and thus were able to inhibit the distracter items from occupying working memory. Individual differences in working memory performance also predict the ability to inhibit automatic responses, such as turning one's eyes toward an item that comes on the screen when the instruction is to turn one's eyes the opposite direction, an 'antisaccade' eye movement.[60]

An alternative view is that differences between high- and low-span individuals in processing efficiency are accompanied by true differences in storage capacity.[5,57] Using a behavioral version of the Vogel et al.[59] distracter task described above, Gold et al.[5] found that the difference between normal adults and those with schizophrenia was mostly in storage capacity. Individuals with schizophrenia ignored distracters about as well as normal adults but still recalled fewer items overall. Additionally, storage capacity, rather than reading speed or other processing skills, accounts for differences in spelling ability among good readers.[54] When carefully teased apart, both storage and the ability to control attention can account for independent variance in measures of intelligence.[57] Individual differences research supports our understanding of working memory as a combination of the limited content held in an easily accessible state, as well as the processing used to encode and maintain that information. Individual variance in each of these components contributes to differences in overall working memory performance and high-level cognitive abilities, such as aptitude test-taking, problem-solving, and reading comprehension. It appears that in general, everyone is limited in what they can remember both by the amount of cognitive space they have to remember items and by how well processing mechanisms can support memory for those items.

Similar to research on within-group individual differences in working memory, we can also use developmental changes to learn about the cognitive functions which underlie working memory.

DEVELOPMENTAL CHANGES IN WORKING MEMORY: EFFICIENCY, CAPACITY, OR BOTH?

Agreement exists that older children and adults demonstrate higher working memory span scores than young children, with adult-level performance on simple tasks attained around age 10.[61] The debate continues, though, over what developmental changes account for this difference. This debate mirrors the one that persists in the individual differences literature, namely whether changes that occur in working memory during childhood involve processing efficiency alone,[60,62] or also include capacity increases during this timeframe.[61]

There are several ways in which the processing abilities of children differ from those of adults. A prominent difference is that young children lack strategies, such as covert rehearsal. Children begin to rehearse, or silently repeat verbal materials, around age seven, but their use of this strategy becomes more complex over time.[63,64] Processing speed also increases throughout childhood and adolescence, with much of the speed increase occurring before the age of 12.[65]

Along with the developing efficiency of processing abilities, the storage component of working memory has also been argued to increase in size throughout development.[57,66] Cowan, Morey, AuBuchon, Zwilling, and Gilchrist[67] showed this in a developmental study similar to Gold et al.'s[5] study of normal and schizophrenic adults. Children were to attend to some items (e.g., colored circles) but ignore others (e.g., colored triangles). Usually they were tested on memory for the color of an attended item but occasionally they were tested on the color of an item they were supposed to ignore. Children in first grade did much better remembering the colors of attended items than ignored items, to the same extent as older children or adults did, provided that there were only four items in the field. Nevertheless, first-grade children remembered far fewer of the objects of either shape, indicating a storage deficit despite adult-like filtering out of irrelevant items in this simple task. Taken together, the studies indicate that processing failures in children are less likely to occur when the amount of information is within the limits of working memory capacity. Moreover, they suggest that age differences in capacity may cause some age differences in processing, rather than depending upon them.

In sum, there is evidence for individual and age group differences in both general working-memory capacity and the ability to manage this capacity using central executive processes. There may also be individual and group differences in other aspects of working memory, such as decay and interference effects.[49]

So far we have relied primarily on behavioral data in examining various claims about working memory. The behavioral data remain important even today but powerful new sources of information are coming from brain imaging studies as well. Some investigators have claimed that brain imaging tells us little about the overall, abstract organization of the cognitive system, but the case for the contribution of brain imaging (i.e., neuroimaging) seems stronger in the area of working memory than in most other areas of research.

WHAT CAN NEUROIMAGING STUDIES TELL US?

The emerging field of neuroimaging holds exciting promise for working memory research, producing new and interesting constraints on theory. Here we give a brief description of findings on working-memory storage and cognitive processing related to the theories and findings discussed above. We attempt to show a correspondence between the behavioral literature and the neuroimaging literature regarding the structure of working memory.

Central to the convergence of behavior-based theory and neuroimaging findings is the observation that items held in working memory tend to activate the same representations in the brain as are activated by perceptual processing. More specifically, the dissociation of spatial and object memory representations in the posterior sections of the brain are upheld in working memory,[68] mirroring the visual perceptual differences of the ventral and dorsal pathways, for what versus where an item is, respectively.[69] Spatial memory storage tends to activate dorsal posterior cortex, whereas object memory tends to activate ventral posterior cortex.[70] Similarly, verbal memory representations show activation in the same areas active in speech perception, and production, primarily the left perisylvian cortex.[71] These findings are consistent with theories positing that items within working memory are composed of activated traces of long-term memory representations.[39–41] It is important to note that although domain-specific representations exist in differing portions of the brain, they still all theoretically could be within a centralized focus of attention, the neural correlate of which may be elsewhere in the brain.

Some researchers have argued that active working memory representations are stored within the frontal cortex.[72] This frontal storage hypothesis was directly tested by Postle and colleagues[73] with the use of repetitive transcranial magnetic stimulation, which temporarily 'turns off' a targeted region of cortex. When areas of the dorsolateral prefrontal cortex (DLPFC) were stimulated, specifically within the middle frontal gyrus, manipulation of items within working memory was disrupted but storage of items was not. In contrast to DLPFC stimulation, when posterior sections of the brain were stimulated, specifically within the superior parietal lobule, both manipulation and storage of items within working memory was disrupted. These results suggest that posterior cortex is necessary for the storage of items within working memory, whereas DLPFC is critical for manipulation but not storage *per se*. It seems parsimonious to assume a common or highly related representational structure between perception and working memory, with the frontal lobes taking the role of the central executive that helps to maintain and manipulate items in storage which are represented by neural activity elsewhere in the brain.

The parietal lobes could be the neural correlate of the focus of attention. Whereas domain-specific traces are represented in various areas used for sensory and conceptual processing, activation in the area of the intra-parietal sulcus (ISP) may underlie the actual chunk capacity limits within working memory. Using a visual working memory task that required memory for the spatial location of colored disks, Todd and Marois[35] demonstrated that the IPS increases in activation as the number of items to be remembered increases, until working-memory capacity is reached, around four items. This cannot be explained as a perceptual effect because as the number of items to be remembered increased beyond four there was no further increase in activation. Xu and Chun[36] confirmed this finding and elaborated on it, demonstrating that inferior IPS activation increases with the number of items to be remembered until it plateaus at 3 or 4 items, whereas superior IPS and lateral occipital complex (LOC) activation reflect the number of items successfully held within working-memory capacity. Thus, superior IPS and LOC activation matched inferior IPS activation when remembering simple items, around four items, but was lower, around two items, when remembering more complex figures.

Together, these studies suggest that the inferior IPS contains information about which items are attended, whereas the superior IPS and LOC index which traces have their features coded and held within working memory. Detailed interpretation of these results is difficult given the current understanding of mechanisms that govern working-memory capacity and performance. Several lines of behavioral research are attempting to address precisely how item complexity effects working memory capacity,[74,75] and seem to indicate that complexity affects the resolution of stimulus representations, whereas the number of representations that can be held is unaffected by item complexity.

Executive functions related to working memory appear to tap differing brain areas, primarily within the frontal cortex. Refreshing processes, such as those theorized by Barrouillet and colleagues to be crucial for working memory maintenance in many situations, appear to result from activation of the DLPFC.[19] Ventrolateral PFC activation, on the other hand, is associated with rehearsal of verbal materials,[19] the critical maintenance mechanism in Baddeley's[8] model working memory. Overall, organization of working memory functions within the brain seem to conform to general patterns of frontal refreshing and manipulation of information and posterior storage of the information, which may include both domain-specific and domain-general types of storage. Theories of working memory based on behavioral data now must be able to explain these and other new findings from brain imaging techniques in addition to data from their traditional paradigms in order to continue to be viable.

CONCLUSION

Research in the past few decades has come a long way towards answering fundamental questions concerning the structure and function of that elusive concept, working memory. For instance, there appears to be a constant capacity limit that can be observed across many experimental paradigms when peripheral sources of storage and maintenance (sensory memory, chunking, and rehearsal) are eliminated. In order to posit multiple independent working-memory systems, one must now be able to explain cross-modal disruption of memory across a wide number of contexts. Research on forgetting has narrowed the range of plausible processes that may be occurring, but has not been able to give a definitive account of how fundamental forgetting from working memory occurs over time. Interference accounts can provide powerful descriptions of most, but not yet all, short-term forgetting. It will be interesting to see if future interference models can maintain their simplicity while at the same time explaining new findings, especially those from cognitive load studies. Firmer conclusions are possible about the reasons for individual and developmental differences in working memory performance. Differences in processing efficiency appear to be responsible for some of the differences in performance across individuals and across different age groups, but differences in working memory storage capacity also appear to account for some of these differences. New questions continue to surface, such as whether an increase in working memory capacity may influence processing efficiency rather than the other way around. Finally, recent research using neuroimaging techniques has made valuable contributions to working memory theory. Working memory storage appears to arise from higher level control functions located in the frontal lobe, and possibly the IPS, maintaining memory traces which consist of perceptual and conceptual features represented in other parts of the brain. This corroborates theoretical descriptions of hierarchically organized models, such as the embedded process model,[11] the TBRS model,[20] and some feature models,[41] although the existing evidence is not specific enough to determine which of these classes of models are the best fit for describing working memory function.

REFERENCES

1. Miller GA, Galanter E Pribram KH. *Plans and the Structure of Behavior*. New York: Holt, Rinehart and Winston; 1960.
2. Baddeley AD, Hitch G. Working memory. In: Bower GH, ed. *The Psychology of Learning and Motivation*, vol. 8. New York: Academic Press; 1974, 47–89.
3. Kyllonen PC, Christal RE. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* 1990, 14:389–433.
4. Martinussen R, Hayden J, Hogg-Johnson S, Tannock R. A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2005, 44:377–384.
5. Gold JM, Fuller RL, Robinson BM McMahon RP Braun EL, et al. Intact attentional control of working memory encoding in schizophrenia. *J Abnorm Psychol* 2006, 115:658–673.
6. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956, 63:81–97.
7. Atkinson RC, Shiffrin RM. Human memory: a proposed system and its control processes. In: Spence KW, Spence JT, eds. *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 2. New York Academic Press; 1968, 89–195.
8. Baddeley AD. *Working Memory*. Oxford Oxford University Press, Clarendon Press; 1986.
9. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 2001, 24:87–185.
10. Baddeley AD. The episodic buffer: a new component of working memory? *Trends Cogn Sci* 2000, 4:417–423.
11. Cowan N. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychol Bull* 1988, 104:163–191.

12. Cocchini G, Logie RH Della Sala S, MacPherson SE, Baddeley AD. Concurrent performance of two memory tasks: evidence for domain-specific working memory systems. *Mem Cognit* 2002, 30:1086–1095.
13. Friedman NP, Miyake A. Differential roles for visuospatial and verbal working memory in situation model construction. *J Exp Psychol Gen* 2000, 129:61–83.
14. Just MA, Carpenter PA, Keller TA, Emery L Zajac H, et al. Interdependence of nonoverlapping cortical systems in dual cognitive tasks. *Neuroimage* 2001, 14:417–426.
15. Morey CC, Cowan N. When do verbal and visual memories conflict? The importance of working memory load and retrieval. *J Exp Psychol Learn Mem Cogn* 2005, 31:703–713.
16. Sirevaag EJ, Kramer AF, Coles MG, Donchin E. Resource reciprocity: an event-related brain potentials analysis. *Acta Psychol* 1989, 70:77–97.
17. Stevanovski B, Jolicoeur P. Visual short-term memory: central capacity limitations in short-term consolidation. *Vis Cogn* 2007, 15:532–563.
18. Saults JS, Cowan N. A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *J Exp Psychol Gen* 2007, 136:663–684.
19. Raye CL, Johnson MK, Mitchell KJ, Greene EJ, Johnson MR. Refreshing: a minimal executive function. *Cortex* 2007, 43:135–145.
20. Barrouillet P, Bernardin S, Camos V. Time constraints and resource sharing in adults' working memory spans. *J Exp Psychol Gen* 2004, 133:83–100.
21. Polich J. Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 2007, 118:2128–2148.
22. Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, et al. The generality of working-memory capacity: a latent-variable approach to verbal and visuo-spatial memory span and reasoning. *J Exp Psychol Gen* 2004, Jun 133:189–217.
23. Chen Z, Cowan N. Core verbal working memory capacity: the limit in words retained without covert articulation. *Q J Exp Psychol* 2009, Jul 62:1420–1429.
24. Broadbent DE. The magic number seven after fifteen years.In: Kennedy A, Wilkes A, eds. *Studies in Long-term Memory*. New York John Wiley & Sons; 1975, 3–18.
25. Cowan N Johnson TD, Saults JS. Capacity limits in list item recognition: evidence from proactive interference. *Memory* 2005, 13:293–299.
26. Halford GS, Maybery MT, Bain JD. Set-size effects in primary memory: an age-related capacity limitation? *Mem Cognit* 1988, 16:480–487.
27. Halford GS, Cowan N, Andrews G. Separating cognitive capacity from knowledge: a new hypothesis. *Trends Cogn Sci* 2007, 11:237–242.
28. Darwin C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London John Murray; 1859.
29. Dirlam DK. Most efficient chunk sizes. *Cogn Psychol* 1972, 3:355–359.
30. MacGregor JN. Short-term memory capacity: limitation or optimization? *Psychol Rev* 1987, 94:107–108.
31. Milner PM. A model for visual shape recognition. *Psychol Rev* 1974, 81:521–535.
32. Gray CM, König P, Engel AK, Singer W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 1989, 338:334–337.
33. Tiitinen H, Sinkkonen J, Reinikainen K, Alho K, Lavikainen J, et al. Selective attention enhances the auditory 40-Hz transient response in humans. *Nature* 1993, 364:59–60.
34. Lisman JE, Idiart MAP. Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science* 1995, 267:1512–1515.
35. Todd JJ Marois R. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 2004, 428:751–754.
36. Xu Y, Chu MM. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 2006, 440:91–95.
37. Peterson LR, Peterson MJ. Short-term retention of individual verbal items. *J Exp Psychol* 1959, 58:193–198.
38. Barrouillet P, Bernardin S, Portrat S, Vergauwe E, Camos V. Time and cognitive load in working memory. *J Exp Psychol Learn Mem Cogn* 2007, 33:570–585.
39. Cowan N. *Attention and Memory: An Integrated Framework*. Oxford Oxford University Press; 1995.
40. Nairne JS. A feature model of immediate memory. *Mem Cognit* 1990, 18:251–269.
41. Oberauer K, Kliegl R. A formal model of capacity limits in working memory. *J Mem Lang* 2006, 55:601–626.
42. Baddeley AD, Thompson N, Buchanan M. Word length and structure of short-term memory. *J Verbal Learn Verbal Behav* 1975, 14:575–589.
43. Cowan N, Day L, Saults JS, Keller TA, Johnson T, et al. The role of verbal output time in the effects of word length on immediate memory. *J Mem Lang* 1992, 31:1–17.
44. Cowan N, Nugent LD, Elliot EM, Geer T. Is there a temporal basis of the word length effect? A response to Service (1998). *Q J Exp Psychol* 2000, 53A:647–660.
45. Mueller ST, Seymour TL, Kieras DE, Meyer DE. Theoretical implications of articulatory duration, phonological similarity and phonological complexity in verbal working memory. *J Exp Psychol Learn Mem Cogn* 2003, 29:1353–1380.

46. Service E. The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Q J Exp Psychol* 1998, 51A:283–304.
47. Lewandowsky S, Oberauer K. The word length effect provides no evidence for decay in short-term memory. *Psychon Bull* 2008, 15:875–888.
48. Cowan N, AuBuchon AM. Short-term memory loss over time without retroactive stimulus interference. *Psychon Bull Rev* 2008, 15:230–235.
49. Cowan N, Elliott EM, Saults JS, Nugent LD, Bomb P, et al. Rethinking speed theories of cognitive development: Increasing the rate of recall without affecting accuracy. *Psychol Sci* 2006, 17:67–73.
50. Lewandowsky S, Duncan M, Brown GDA. Time does not cause forgetting in short term serial recall. *Psychon Bull Rev* 2004, 11:771–790.
51. Oberauer K, Lewandowsky S. Forgetting in immediate serial recall: decay, temporal distinctiveness, or interference? *Psychol Rev* 2008, 115:544–576.
52. Portrat S, Barrouillet P, Camos V. Time-related decay or interference-based forgetting in working memory? *J Exp Psychol Learn Mem Cogn* 2008, 34:1561–1564.
53. Daneman M, Carpenter PA. Individual differences in working memory and reading. *J Verbal Learn Verbal Behav* 1980, 19:450–466.
54. Ormrod JE, Cochran KF. Relationship of verbal ability and working memory to spelling achievement and learning to spell. *Read Res Instruction* 1988, 28:33–43.
55. Engle WE, Tuholski SW, Laughlin JE, Conway AR. Working memory, short-term memory, and general fluid intelligence: a latent variable approach. *J Exp Psychol Gen* 1999, 128:309–331.
56. McCollough AW, Vogel EK. Your inner spam filter: What makes you so smart? Might be your lizard brain. *Scient Am Mind* 2008, 32–35.
57. Cowan N, Fristoe NM, Elliot EM, Brunner RP, Saults JS. Scope of attention, control of attention, and intelligence in children and adults. *Mem Cognit* 2006, 34:1754–1768.
58. Conway ARA, Cowan N, Bunting MF. The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychon Bull Rev* 2001, 8:331–335.
59. Vogel EK, McCollough AW, Machizawa MG. Neural measures reveal individual differences in controlling access to working memory. *Nature* 2005, 438:500–503.
60. Kane MJ, Bleckley MK, Conway ARA, Engle RW. A controlled-attention view of working-memory capacity. *J Exp Psychol Gen* 2001, 130:169–183.
61. Cowan N, Elliott EM, Saults JS, Morey CC, Mattox S, et al. On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cogn Psychol* 2005, 51:42–100.
62. Case R, Kurland DM, Goldberg J. Operational efficiency and the growth of short-term memory span. *J Exp Child Psychol* 1982, 33:386–404.
63. Cowan N, Saults JS, Morey CC. Development of working memory for verbal-spatial associations. *J Mem Lang* 2006, 55:274–289.
64. Gathercole SE, Adams AM, Hitch GJ. Do young children rehearse? An individual-differences analysis. *Mem Cognit* 1994, 22:201–207.
65. Fry AF, Hale S. Processing speed, working memory, and fluid intelligence. *Psychol Sci* 1996, 7:237–241.
66. Cowan N, Nugent LD, Elliot EM, Ponomarev I, Saults JS. The role of attention in the development of short-term memory: age differences in the verbal span of apprehension. *Child Dev* 1999, 70:1082–1097.
67. Cowan N, Morey CC, AuBuchon AM, Zwilling CE, Gilchrist AL. Seven year-olds allocate attention like adults do unless working memory is overloaded. *Dev Sci* 2009, 12:1–14.
68. Jonides J, Lacey SC, Nee DE. Processes of working memory in mind and brain. *Curr Dir Psychol Sci* 2005, 14:2–5.
69. Haxby JV, Grady CL, Horwitz B, Ungerleider LG, Mishkin M, et al. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc Natl Acad Sci USA* 1991, 88:1621–1625.
70. Postle BR, D'Esposito M. "What"-then-"where" in visual working memory: an event-related fMRI study. *J Cogn Neurosci* 1999, 11:585–597.
71. Feredoes E, Tononi G, Postle BR. The neural bases of short-term storage of verbal information are anatomically variable across individuals. *J Neurosci* 2007, 27:11003–11008.
72. Leung HC, Gore JC, Goldman-Rakic PS. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *J Cogn Neurosci* 2002, 14:659–671.
73. Postle BR, Ferrarelli F, Hamidi M, Feredoes E, Massimini M, et al. Repetitive transcranial magnetic stimulation dissociates working memory manipulation from retention functions in the prefrontal, but not posterior parietal, cortex. *J Cogn Neurosci* 2006, 18:1712–1722.
74. Awh E, Barton B, Vogel EK. Visual working memory represents a fixed number of items regardless of complexity. *Psychol Sci* 2007, 18:622–628.
75. Eng HY, Chen D, Jiang Y. Visual working memory for simple and complex visual stimuli. *Psychon Bull Rev* 2005, 12:1127–1133.

FURTHER READING

Baddeley A. *Working Memory, Thought, and Action*. New York: Oxford University Press; 2007.

Conway ARA, Jarrold C, Kane MJ, Miyake A, Towse, J, eds. *Variation in Working Memory*. New York: Oxford University Press; 2007.

Cowan N. *Working Memory Capacity*. Hove, East Sussex, UK: Psychology Press; 2005.

Jonides J, Lewis RL, Nee DE, Lustig CA, Berman MG, etal. The mind and brain of short-term memory. *Ann Rev Psychol* 2008, 59: 193–224.

Klingberg T. *The Overflowing Brain: Information Overload and the Limits of Working Memory*. New York: Oxford University Press; 2009. [Translated by Neil Betteridge.]

Osaka N, Logie RH, D'Esposito M. *The Cognitive Neuroscience of Working Memory*. New York: Oxford University Press; 2007.