This chapter was originally published in the book The Psychology of Learning and Motivation, Volume 65. The copy attached is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research, and educational use. This includes without limitation use in instruction at your institution, distribution to specific colleagues, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution’s website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier’s permissions site at:
http://www.elsevier.com/locate/permissionusematerial

ISBN: 9780128047903
Copyright © 2016 Elsevier Inc. All rights reserved.
Academic Press
CHAPTER ONE

The Many Facets of Individual Differences in Working Memory Capacity

Nash Unsworth
University of Oregon, Eugene, OR, United States
E-mail: nashu@uoregon.edu

Contents
1. Introduction 2
2. Importance of Working Memory 2
3. A Theoretical Framework for Working Memory Capacity 5
4. Multiple Facets Influence Individual Differences in Working Memory Capacity 7
   4.1 Capacity of Primary Memory 7
   4.2 Attention Control 16
   4.3 Secondary Memory 25
5. Measurement of Working Memory Capacity 32
6. Heterogeneity of Working Memory Capacity Limitations 36
7. Conclusions 37
References 37

Abstract
This chapter reviews prior research and our current thinking on individual differences in working memory capacity (WMC), the nature of WMC limitations, and the relation between WMC and higher-order cognition (in particular fluid intelligence). Evidence is reviewed suggesting that individual differences in WMC arise from multiple different facets. These facets include differences in the capacity of primary memory, attention control abilities, and secondary memory abilities. We review evidence suggesting that each facet is related to overall individual differences in WMC and part of the reason for the predictive power of WMC. Furthermore, we outline the role of each facet in various measures of WMC including complex span tasks, simple span tasks, and visual arrays change detection tasks. We argue that to understand WMC and individual differences in WMC, we must delineate and understand the various facets that make up WMC.
1. INTRODUCTION

Researchers interested in both experimental and differential psychology have long argued for the need to include individual differences in theory construction (Cohen, 1994; Cronbach, 1957; Kosslyn et al., 2002; Melton, 1967; Underwood, 1975). In particular, it has been suggested that theories of memory and attentional processes (and cognition in general) need to attempt to account for individual differences in the ability to carry out the processes specified in the theory. Although interest in individual differences in cognitive processes has waxed and waned over the years, one area that has seen fairly continual interest is that of immediate memory processes. This chapter reviews prior research and our current thinking on individual differences in working memory capacity (WMC), the nature of WMC limitations, the role of WMC in cognitive tasks, and the relation between WMC and higher-order cognition. Although there are many other excellent research programs studying working memory and individual differences in WMC, here we primarily focus on our own work. As will be seen, our work draws on prior reviews published in this series including Atkinson and Shiffrin (1968), Baddeley and Hitch (1974), Cowan, Morey, Chen, Gilchrist, and Saults (2008), and Engle and Kane (2004), among others. Like prior calls to combine experimental and differential methods, we use individual differences as a means of not only understanding differences among individuals in cognitive capabilities, but also to better understand the nature and function of working memory more broadly.

2. IMPORTANCE OF WORKING MEMORY

Research examining immediate memory is typically cast in frameworks distinguishing information that is utilized over the short-term from information that is utilized over the long-term. Initially, immediate memory was conceptualized as a somewhat passive repository of information before that information was transferred to long-term or secondary memory. In early modal models of memory, immediate memory was seen as having limited capacity and important task-relevant information was maintained primarily via verbal rehearsal. If the information was not rehearsed, then it was rapidly lost from the system.

Despite the importance of immediate memory and a wealth of data supporting a division between immediate and long-term memory, it soon
became clear that immediate memory, as initially conceptualized, was overly simplistic in terms of being a simple passive buffer. With this limitation clearly in mind Atkinson and Shiffrin (1971) and Baddeley and Hitch (1974), among others, argued for a dynamic memory system where the function of immediate memory was to carry out cognitive operations important for a wide variety of tasks. Specifically, Baddeley and Hitch (1974) argued for a memory system that could simultaneously manipulate the current contents of memory as well as update information in memory to accomplish task goals. They called this system working memory to emphasize the need for actively working with information rather than simply passively holding onto the information (see also Atkinson & Shiffrin, 1968, 1971; Miller, Galanter, & Pribram, 1960). Early prominent models of working memory suggested that it was not only a system responsible for actively maintaining task-relevant information, but also a system composed of many important control processes that ensure proper maintenance, storage, and retrieval of that information (e.g., Atkinson & Shiffrin, 1968, 1971; Baddeley & Hitch, 1974). These control processes included rehearsal, coding, organization, and retrieval strategies. Importantly, these control processes were thought to be needed for coordinating the many subcomponent processes necessary for processing new information and to retrieve relevant old information. This conceptualization placed working memory at the forefront of explaining complex cognitive activities.

Given the theoretical importance of working memory in a broad array of tasks and situations, research over the last 35 plus years has been aimed at examining the predictive power of working memory. That is, the capacity of working memory should be related to a number of measures that rely on working memory. Largely beginning with Daneman and Carpenter (1980) research has found that individual differences in WMC are one of the best predictors of a broad array of cognitive capabilities. Specifically, research has shown that measures of WMC are related to reading and language comprehension (Daneman & Carpenter, 1980), complex learning (Kyllonen & Stephens, 1990), performance on standardized achievement tests (Engle, Tuholski, Laughlin, & Conway, 1999), and vocabulary learning (Daneman & Green, 1986). Thus, as theorized, measures of WMC demonstrate strong and consistent relations with a broad array of cognitive abilities that are thought to rely on working memory processes.

Beginning with the work of Kyllonen and Christal (1990) research has suggested that there is a strong link between individual differences in WMC and intelligence (see also Engle et al., 1999; Kane et al., 2004). In
particular, this work suggests that at an individual task level, measures of WMC correlate with fluid intelligence (gF) around 0.45 (Ackerman, Beier, & Boyle, 2005) and at the latent level, WMC and gF are correlated around 0.72 (Kane, Hambrick, & Conway, 2005). Thus, at a latent level WMC and gF seem to share approximately half of their variance. As a further example of this relation, we reanalyzed data from 867 participants from our laboratory each of which had completed three WMC measures and three gF measures. Shown in Fig. 1 is the resulting latent variable model. As can be seen, WMC and gF abilities were strongly related. These examples demonstrate that WMC and gF are strongly related and share a good deal of common variance. Furthermore, these results demonstrate that this important relation is domain-general in nature given that both the WMC and gF factors were made up by tasks varying in their content. This suggests that whatever the reasons for the relation between WMC and fluid abilities, they are likely domain-general and cut across multiple different types of tasks.

Additionally, not only has WMC been implicated in higher-order cognition, but WMC is also implicated in other research domains. For example, measures of WMC predict early onset Alzheimer’s disease (Rosen, Bergeson, Putnam, Harwell, & Sunderland, 2002), one’s ability to deal with life-event stress (Klein & Boals, 2001), aspects of personality (Unsworth

![Figure 1](image_url)  
**Figure 1** Confirmatory factor analysis for working memory capacity (WMC) and fluid intelligence (gF). Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Raven = Raven Advanced Progressive Matrices; LS = letter sets; NS = number series. All paths and loadings are significant at the $p < 0.05$ level.
et al., 2009), susceptibility to choking under pressure (Beilock & Carr, 2005), and stereotype threat (Schamader & Johns, 2003). Furthermore, various neuropsychological disorders, including certain aphasias (Caspari, Parkinson, LaPointe, & Katz, 1998), Alzheimer’s disease (Kempler, Almor, Tyler, Andersen, & MacDonald, 1998), schizophrenia (Stone, Gabrieli, Stebbins, & Sullivan, 1998), and Parkinson’s disease (Gabrieli, Singh, Stebbins, & Goetz, 1996), have been linked to deficits in WMC. Thus, the utility of WMC is not merely limited to performance on high-level cognitive tasks, but is also important in a variety of situations that impact people on a day-to-day basis.

3. A THEORETICAL FRAMEWORK FOR WORKING MEMORY CAPACITY

Based on prior work we have developed a theory of individual differences in WMC which suggests that individual differences in WMC result from multiple facets, each of which is important for performance on a variety of tasks (Unsworth, 2014; Unsworth & Engle, 2007; Unsworth, Fukuda, Awh, & Vogel, 2014; Unsworth & Spillers, 2010a). Similar to prior conceptions, we think of working memory as consisting of memory units active above some threshold that can be represented via a variety of different codes (phonological, visuospatial, semantic, etc.), as well as a set of general purpose control processes (e.g., Atkinson & Shiffrin, 1971; Cowan, 1988; 1995). Specifically, in line with classic dual-component models of memory, we suggest that there is a limited capacity component important for maintaining information over short time intervals and a larger more durable component important for maintaining information over longer time intervals (Atkinson & Shiffrin, 1968; Raaijmakers & Shiffrin, 1980). Similar to James (1890), we refer to these two components as primary memory (PM) and secondary memory (SM; c.f. Craik, 1971; Craik & Levy, 1976). Thus, similar to the model initially proposed by Atkinson and Shiffrin (1971), working memory represents both the activated portion of the long-term repository and the set of control processes that act on those activated representations to bring them into a heightened state of activation and actively maintain them in the face of distraction (see also Engle et al., 1999).

In this framework, attention control processes serve to actively maintain a few distinct representations for online processing in PM. These representations include things such as goal states for the current task, action plans, partial solutions to reasoning problems, and item representations in list
memory tasks. In this view, as long as attention is allocated to these representations, they will be actively maintained in PM (Craik & Levy, 1976). This continued allocation of attention serves to protect these representations from interfering internal and external distraction (eg, Engle & Kane, 2004; Unsworth & Engle, 2007). However, if attention is removed from the representations due to internal or external distraction or due to the processing of incoming information that exceed capacity, these representations will no longer be actively maintained in PM and therefore, will have to be retrieved from SM if needed. Accordingly, SM relies on a cue-dependent search mechanism to retrieve items (Raaijmakers & Shiffrin, 1980; Shiffrin, 1970). Additionally, the extent to which items can be retrieved from SM will be dependent on overall encoding abilities, the ability to reinstate the encoding context at retrieval, and the ability to focus the search on target items and exclude interfering items (ie, proactive interference). Similar to Atkinson and Shiffrin (1968, 1971) this framework suggests that working memory is not only a state of activation, but also represents the set of control processes that are needed to maintain that state of activation, to prevent other items from gaining access to this state of activation, and to bring other items into this state of activation via controlled retrieval (Engle et al., 1999). Thus, working memory represents a dynamic interface between information present in the environment and our repository of past experiences.

Within the current framework, individual differences in WMC arise from multiple different factors. Specifically, as discussed more thoroughly throughout, individual differences in WMC arise from differences in the capacity of PM, differences in attention control processes that serve to maintain task-relevant information in PM, and differences in control processes that ensure that task-relevant information is properly encoded in and retrieved from SM. Thus, we will suggest that there are three primary reasons for differences in WMC, and each of these different facets is important for the predictive power of WMC. That is, measures of WMC are related to performance in a wide variety of tasks and situations. It seems unlikely that there is a single cause/mechanism responsible for these relations. Indeed, prior research has consistently shown that if you covary out one primary cause (such as attention control) the relation between WMC and some other variable (eg, gF) is reduced but not completely eliminated (ie, Unsworth, 2014; Unsworth & Spillers, 2010a). Thus, it is unlikely that individual differences in WMC reduce to a single common cause. Here we suggest that WMC represents a number of important related facets, each of which is important for higher-order cognitive processes. Furthermore, we suggest
that individuals may differ on some, or all of these facets, thereby determining the relation with other measures. Collectively, this suggests that there are multiple functional roles that WMC plays, and points to the multifaceted nature of individual differences in WMC. In the next sections, we discuss in detail ours and related work on these facets.

4. MULTIPLE FACETS INFLUENCE INDIVIDUAL DIFFERENCES IN WORKING MEMORY CAPACITY

4.1 Capacity of Primary Memory

We consider PM as the small set of items that are in heightened state of activation and the current focus of processing. That is, the small set of items that an individual is currently consciously working with. We have argued that the function of PM is to maintain a distinct number of separate representations active for ongoing processing. These representations remain active via the continued allocation of attention. This is consistent with prior work by Craik and Levy (1976) who suggested that “the capacity of primary memory is the number of events that can be attended to simultaneously or the number of internal representations that can be simultaneously activated by the process of attention” (Craik & Levy, 1976, p. 166). Thus, PM is the small set of items that are being maintained in mind from the environment or the small set of items that are reactivated from our long-term repository. Craik and Levy (1976) go on to note that “information is ‘in PM’ only by virtue of the continued allocation of attention; when attention is diverted the trace is left in SM” (p. 166). Similar to Craik and Levy (1976) we assume that an item is in PM if it is currently be attended to. If attention is directed elsewhere, due to processing new information or having attention captured by internal (mind-wandering) or external distraction, representations will be displaced from PM. Similar to the view advocated here, Craik and Levy (1976) argued that the capacity of PM is the capacity to maintain a distinct number of representations by continually paying attention to those representations. This suggests that a key aspect to PM is the ability to individuate and apprehend multiple items and maintain those items in an active state to facilitate the further processing of task-relevant information (Cowan, 2001).

PM is also thought to be a highly flexible component that changes depending on the current context and goals (Atkinson & Shiffrin, 1968, 1971; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005). That is, PM is not simply a buffer limited to a particular number
of slots, but rather is a more dynamic system that can change due to task demands. In particular, in tasks and situations where many representations need to be maintained (such as remembering a long list of items), the capacity of PM will be maximal. This is because at recall, items that are in PM are simply unloaded and recall is nearly perfect. Furthermore, maintaining items in PM selectively protects those items from proactive interference (PI; Craik & Birtwistle, 1971; Unsworth & Engle, 2007; Wickens, Moody, & Dow, 1981). In other tasks where only a single important representation needs to be maintained (such as maintaining an important goal), the capacity of PM will shrink to encapsulate only this one representation. In both situations, the representations are maintained by continually paying attention to them. If attention is captured by distracting external or internal stimuli, the information will fail to be actively maintained leading to decrements in performance.

Based on a great deal of evidence, PM is thought to have a capacity of approximately 4 ± 1 items (Broadbent, 1975; Cowan, 2001). When more than four items are present, items currently within PM are probabilistically displaced and must be recalled from SM. Evidence for a four-item limit comes from a variety of behavioral and physiological studies. For example, Cowan (2001) (see also Cowan et al., 2008) reviewed a wealth of evidence from the prior reviews of Broadbent (1975) and Watkins (1974) as well as much more recent evidence from a number of tasks and found that the average capacity was close to four items. For example, estimates of visual working memory obtained from visual arrays tasks suggest a capacity of approximately four items (Luck & Vogel, 1997). Similar estimates arise when examining multiobject tracking, the influence of proactive interference on recall, the subitizing range, and parameter estimates of capacity in mathematical models of memory and cognition. In nearly all cases four or so items seemed to be maintained. Cowan (2001) suggested that capacity of the focus of attention (or PM) was roughly four items. Additionally, it should be noted that similar estimates are obtained when using a variety of materials and variety of presentation modes suggesting that PM is a domain-general system that maintains a distinct set of items regardless of their particular code (Li, Christ, & Cowan, 2014).

Recent neural and physiological evidence corroborates the behavioral estimates of capacity. For example, using functional magnetic resonance imaging (fMRI), Todd and Marois (2004) found that the delay signal in the intraparietal sulcus increased as set size increased, reaching asymptote around three to four items. Examining event-related potentials, Vogel and
Machizawa (2004) demonstrated that sustained activity over posterior parietal electrodes during the delay of a visual working memory task increased as set size increased and reached asymptote around three to four items. This activity, known as the contralateral delay activity (CDA), reflects a sustained negative wave at posterior electrodes contralateral to the attended hemifield. Importantly, the CDA seems to track the number of items currently being maintained in PM (Vogel & Machizawa, 2004).

Recently, we examined whether phasic pupillary responses would also track the number of items being maintained in PM over a brief delay (Unsworth & Robison, 2015a). Much prior research has shown that the pupil dilates in response to the cognitive demands of a task (Beatty, 1982). For example, Kahneman and Beatty (1966) demonstrated that pupillary dilation increased as more items were required for recall in a standard short-term memory task (see also Peavler, 1974). These effects reflect task-evoked phasic pupillary responses in which the pupil dilates relative to baseline levels due to increases in cognitive processing load. A number of studies have demonstrated similar phasic pupillary responses in a variety of tasks (Beatty & Lucero-Wagoner, 2000). These and other results led Kahneman (1973) and Beatty (1982) to suggest that phasic pupillary responses correspond to the intensive aspect of attention and provide an online indication of the utilization of capacity (see also Beatty & Lucero-Wagoner, 2000). Thus, assuming that PM capacity reflects the number of items that can be maintained via the continued allocation of attention, we should see that attention is allocated to items during the delay to maintain them in PM, and that as the amount of information that needs to be maintained increases, so should the amount of attentional allocation. Importantly, this increase in attention allocation should increase only up to capacity limits, at which point no more attention can be allocated resulting in leveling off. To examine this, we had participants perform a visual arrays change detection task in which the number of items to be maintained varied from one to eight and participants’ pupils were measured continuously throughout the task. Consistent with prior research, behavioral PM capacity was estimated at close to four items (Cowan, 2001). Importantly, phasic pupillary responses increased as set size increased and then plateaued between around four items consistent with the behavioral estimate of PM capacity. Additionally, the phasic response maintained throughout the delay period suggesting that participants were continuously allocating effortful attention to the items to actively maintain them in PM. Collectively, these results suggest that the capacity of PM is limited to four or so items and this capacity limit, results from the fact
that only four or so items can be distinctly maintained via the continued allocation of attention.

In terms of individual differences in WMC, we and others (e.g., Cowan et al., 2005; Cowan, Fristoe, Elliot, Brunner, 2006) have suggested that a critical determinant is the number of items that can be maintained in PM. That is, individual differences in the capacity of PM is one of the main sources of variance contributing to individual differences in WMC, and part of the reason WMC relates to higher-order cognitive constructs like gF. Based on prior work by Broadbent (1975) and Cowan (2001) there are three main ways in which individual differences in PM capacity have been assessed. Although there are a number of different ways of assessing PM capacity, these three have been used most frequently. These include obtaining estimates of PM capacity from immediate free recall, estimating capacity from errorless performance on simple span tasks, and estimating capacity from visual arrays change detection tasks. Each of these has been shown to demonstrate substantial individual differences, and each has been shown to correlate with measures of WMC and gF. For example, consider PM estimates obtained from immediate free recall. Here participants are given a list of items (typically words), and after the last word participants are instructed to recall all of the items they can in any order they wish. A number of methods have been developed in an attempt to estimate the contributions of PM and SM in these tasks (e.g., Watkins, 1974). In prior research we and others have relied on Tulving and Colotla's (1970) method. In this method, the number of words between a given word's presentation and recall was tallied. If there were seven or fewer words intervening between presentation and recall of a given word, the word was considered to be recalled from PM. If more than seven words intervened, then the word was considered to be recalled from SM. This method suggests that items in PM are those items that are recalled first, with only a minimal amount of interference from input and output events (Watkins, 1974). Importantly, this method does not suggest that all recency items are recalled from PM, rather only those recency items that are recalled first. It is entirely possible that participants will recall a recency item after many other items have been recalled, in which case that item would be considered to be recalled from SM. Prior work has suggested that this method provides fairly valid estimates of PM and SM (Watkins, 1974). With this method we have repeatedly shown that high WMC individuals have higher estimates of PM capacity than low WMC individuals (see Fig. 2). Furthermore, these estimates correlate well with measures of WMC and with measures of
intelligence (e.g., Engle et al., 1999; Unsworth, Spillers, & Brewer, 2010; Shipstead, Lindsey, Marshall, & Engle, 2014).

Similar results are obtained when estimating PM capacity via errorless performance in simple span tasks. Specifically, as suggested by Broadbent (1975), one can estimate PM capacity by examining the point at which participants drop off of perfect performance on simple span tasks. Using this method we (Unsworth & Engle, 2006) found that estimates of PM capacity were around four items and that these estimates correlated with WMC and gF. Similar to the results obtained with immediate free recall, high WMC individuals have larger estimates of PM capacity than low WMC individuals (see Fig. 2). To see if these results replicate, we reanalyzed data from Engle et al. (1999) examining errorless performance (see Unsworth, 2014). As shown in Fig. 2, similar differences in PM capacity between high and low WMC individuals were found. Furthermore, as shown in Fig. 3, when examining performance as a function of list-length, it is clear that performance is very high for short list-lengths. For larger list-lengths there is a large drop in performance, and this drop in performance occurs earlier for low WMC individuals than for high WMC individuals. Importantly, we also examined the extent to which estimates of PM capacity from immediate free recall and errorless performance on simple span tasks would correlate and load on the same factor. Shown in Fig. 4A is a confirmatory factor

Figure 2 Estimates of primary memory capacity for high and low working memory individuals on immediate free recall (IFR), errorless performance on simple span tasks (SS), and change detection (CD). IFR1 is from Unsworth and Engle (2007); IFR2 is from Engle et al. (1999); IFR3 is from Unsworth, Spillers, et al. (2010); SS1 from Engle et al. (1999) (reanalyzed by Unsworth, 2014); SS2 is from Unsworth and Engle (2006); CD is from Unsworth et al. (2014).
analysis demonstrating that estimates of PM capacity from the different methods correlate and load on the same latent factor. Importantly, this latent factor is related to both WMC and gF. Thus, similar estimates are obtained from the different methods, and these capacity estimates are related to individual differences in WMC and gF.

Another method for estimating PM capacity prominently used in studies of visual working memory comes from visual arrays change detection tasks. In this task, participants are briefly shown an array of items (such as colored squares) and following a brief delay are presented with a test array in which one of the items may have changed colors. The participant’s task is to indicate if one of the items has changed color or not (Luck & Vogel, 1997). Similar to examining errorless performance on simple span tasks, prior research has shown that performance is good up until around four items, after which performance gets steadily worse (Luck & Vogel, 1997). Using a formula to estimate capacity in these tasks has shown that capacity (k) is typically around three to four items with substantial individual differences. Importantly, variance in capacity from these tasks is related to other measures of WMC such that high WMC individuals have larger capacities than low WMC individuals (see Fig. 2). Additionally, a number of recent studies have found that individual differences in capacity in these tasks is related to higher-order cognition and are part of the reason why WMC is related

![Figure 3 Proportion correct as a function of list-length in simple span tasks for high and low working memory capacity (WMC) individuals. Data is from Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: evidence from list-length effects. Journal of Memory and Language, 54, 68–80.](image-url)
to higher-order cognition (e.g., Cowan et al., 2005, 2006; Fukuda, Vogel, Mayr, & Awh, 2010; Shipstead, Redick, Hicks, & Engle, 2012, 2014; Unsworth et al., 2014). For example, shown in Fig. 4B is a reanalysis of Shipstead et al. (2014) in which measures of PM capacity from immediate free recall and the change detection tasks are allowed to load on the same latent factor, and this factor is allowed to correlate with factors for WMC and gF. As can be seen, capacity estimates from the two methods correlate and load with similar magnitudes on the PM factor. Importantly, this factor is strongly related to the WMC and gF factors. Thus, the variance in common between PM estimates from immediate free recall and change detection index is an important individual difference that is related to WMC and gF. We suggest that this shared variance is an index of an individual’s ability to actively maintain distinct pieces of information in PM, regardless of the nature or modality of that information. That is, what is shared across the verbal (immediate free recall) and visual (change detection) estimates of PM capacity is a critical reason for individual differences in WMC.

In addition to demonstrating individual differences in behavioral estimates of capacity, a number of recent studies have found physiological correlates of PM capacity as well. For example, as mentioned previously, Todd and Marois (2004) found that activity in the intraparietal sulcus asymptoted around three to four items. Importantly in a subsequent study Todd and Marois (2005) found that the delay activity predicted individual differences in behavioral estimates of working memory capacity. Furthermore, Vogel and Machizawa (2004) demonstrated that the CDA not only plateaued around three to five items, but it was also strongly related to behavioral estimates of an individual’s capacity. A number of subsequent studies have shown that the CDA provides an index of an individual’s capacity. Indeed, in a recent latent variable study we (Unsworth, Fukuda, Awh, & Vogel, 2015) found that the CDA across different tasks correlated ($r = 0.65$) and loaded on the same latent factor. Importantly, this latent CDA factor was related to behavioral estimates of capacity ($r = -0.37$), as well as latent factors of WMC ($r = -0.20$) and gF ($r = -0.49$). Thus, neural markers of PM capacity are potent predictors of individual differences in WMC and higher-order cognition.

Another physiological correlate of PM capacity is pupil diameter. Earlier we described a study where we examined pupillary correlates of PM capacity, demonstrating that phasic pupillary responses during a delay in a change detection task increased until around four items and then plateaued...
(A) Confirmatory factor analysis for working memory capacity (WMC), fluid intelligence (gF), and primary memory (PM) with PM estimates from immediate free recall and errorless performance in two simple span tasks. Ospan = operation span;
In that study we also examined individual differences. We found that behavioral estimates of capacity correlated with phasic pupillary responses ($r = 0.43$), suggesting that high WMC individuals were able to maintain more items in PM than low WMC individuals due to a greater allocation of attention. Furthermore, assuming that actively maintaining items throughout a delay is effortful, we should see an increase in pupil diameter at the beginning of the delay, this increase should be maintained throughout the delay, and this should differ between high and low WMC individuals. This is precisely what was found. For example, shown in Fig. 5 are the phasic pupillary responses (set sizes four to eight averaged together) for high and low WMC individuals. For high WMC individuals there is a sharp increase early in the delay period and this maintains throughout the delay. For low WMC individuals the increase is more gradual throughout the delay period, and low WMC individuals do not quite reach the same level as high WMC individuals. This suggests that when presented with a number of items that meet or exceed one’s capacity, effortful attention is needed to maintain those items throughout a delay, and high WMC individuals are better able to allocate attention to those items than low WMC individuals.

Estimates of capacity from various sources (different tasks, physiological and neural markers) share considerable variance and seem to reflect a common ability. We and others suggest that the capacity of PM reflects the ability to maintain a few important and task-relevant representations in a highly active state for ongoing processing. These representations are maintained via the continued allocation of attention, and there are substantial individual differences in this capacity. Variability in PM capacity is a critical reason for individual differences in WMC and a main reason why

Cspan = counting span; Rspan = reading span; Raven = Raven Progressive Matrices; Cattell = Cattell’s Culture Fair Test; IFRPM = primary memory estimate from immediate free recall; FDPM = primary memory estimate from forward span with phonologically dissimilar words; FSPM = primary memory estimate from forward span with phonologically similar words. All paths and loadings are significant at the $p < 0.05$ level. (B) Confirmatory factor analysis for WMC, gF, and PM with PM estimates from immediate free recall and $k$ estimates from change detection. Ospan = operation span; Sym-span = symmetry span; Raven = Raven Advanced Progressive Matrices; LS = letter sets; NS = number series; IFRPM1 = primary memory estimate from immediate free recall; IFRPM2 = primary memory estimate from immediate free recall; CDPM2 = primary memory estimate from change detection; CDPM2 = primary memory estimate from change detection.
measures of WMC correlate so well with measures of higher-order cognition (particularly gF).

4.2 Attention Control

We consider attention control (AC) as the set of attentional processes that aid in the ability to actively maintain information in PM in the presence of interference and distraction. That is, AC refers to the ability to select and actively maintain items in the presence of internal and external distraction (Engle & Kane, 2004). In particular, AC abilities are necessary when goal-relevant information must be maintained in a highly active state in the presence of potent internal and external distraction. Any lapse of attention (or goal neglect, Duncan, 1995; De Jong, Berendsen, & Cools, 1999) will likely lead to a loss of the task goal and will result in attention being automatically captured by internal (eg, mind-wandering; Kane et al., 2007; McVay & Kane, 2012a) or external distraction (eg, Fukuda & Vogel, 2009; Unsworth et al., 2014; Unsworth & McMillan, 2014a). Thus, AC abilities are needed to protect items that are being held in PM, to effectively select target representations for active maintenance, to filter out irrelevant distractors and prevent them from gaining access to PM (eg, Vogel, McCollough, & Machizawa, 2005), and to sustain a consistent level of attention across trials.

As a classic example, consider the antisaccade task (Hallet, 1978). In this task, participants must direct their gaze and their attention either toward (prosaccade) or away (antisaccade) from a flashing cue. On prosaccade trials,
the task goal and the prepotent response coincide (eg, look at the flashing box). Relying on either goal maintenance or automatic orienting will result in the correct behavior. On antisaccade trials, however, the task goal and the prepotent response conflict (eg, if flashing on left, look right). Thus, on antisaccade trials it is critically important to maintain the task goal in order for accurate responding to occur. If the task goal is not actively maintained, any momentary lapse in attention will result in attentional capture by the cue (Roberts, Hager, & Heron, 1994; Roberts & Pennington, 1996). Thus, any lapses in attention will result in the prepotent response guiding behavior and the occurrence of a fast reflexive error (ie, looking at the flashing cue), or a much slower than normal response time. In terms of individual differences, high and low WMC individuals differ in the extent to which they can maintain representations in an active state, including goal representations, and thus low WMC individuals should demonstrate poorer performance on antisaccade trials which is exactly the case (Kane, Bleckley, Conway, & Engle, 2001; Unsworth, Schrock, & Engle, 2004; Unsworth, Redick, et al., 2012). Specifically, low WMC individuals make more antisaccade errors (ie, they are more likely to look at the flashing cue) and have slower correct reaction times than high WMC individuals suggesting that they are more susceptible to goal neglect. Indeed, reanalyzing data from 1038 participants in our laboratory suggests that WMC and antisaccade accuracy are consistently correlated ($r = 0.31$). Thus, a key aspect of AC is the ability to actively maintain the current goal in a highly active state and prevent attentional capture.

These AC abilities are needed in a host of tasks which have been shown to correlate with WMC. For example, in addition to antisaccade, WMC differences have been demonstrated in Stroop interference (Kane & Engle, 2003; Meier & Kane, 2013; Morey et al., 2012), flanker interference (Heitz & Engle, 2007; Redick & Engle, 2006), dichotic listening (Colflesh & Conway, 2007; Conway, Cowan, & Bunting, 2001), performance on the psychomotor vigilance task (Unsworth, Redick, et al., 2010; Unsworth & Spillers, 2010a), performance on the Sustained Attention to Response Task (SART; McVay & Kane, 2009), performance on versions of go/no-go tasks (Redick, Calvo, Gay, & Engle, 2011), performance on the AX-CPT task (Redick, 2014; Redick & Engle, 2011; Richmond, Redick, & Braver, 2015), performance on cued visual search tasks (Poole & Kane, 2009), performance on attentional capture tasks (Fukuda & Vogel, 2009, 2011), and performance on some versions of the Simon task (Meier & Kane, 2015).
Figure 6 (A) Confirmatory factor analysis for working memory capacity (WMC), fluid intelligence (gF), and attention control (AC). Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Raven = Raven Advanced Progressive Matrices; LS = letter sets; NS = number series; Anti = antisaccade; Flanker = flanker interference score; PVT = psychomotor vigilance task. All paths and loadings are significant at the
Across a number of studies, individual differences in WMC have been shown to be related to performance on a number of AC tasks. These differences are found not only when examining individual AC measures, but also when examining latent variables composed of the shared variance among multiple AC tasks. For example, Unsworth and Spillers (2010) had participants perform a number of WMC tasks as well as antisaccade, flankers, Stroop, and the psychomotor vigilance task as measures of AC. We found that all of the AC tasks loaded on the same AC factor and this factor was strongly related to latent WMC and gF factors (see also McVay & Kane, 2012; Unsworth et al., 2014; Unsworth & McMillan, 2014a). Indeed, as a further demonstration of the robustness of the AC relation with WMC and gF, shown in Fig. 6A is a confirmatory factor analysis examining data from 646 participants in our laboratory. As can be seen, antisaccade accuracy, flanker interference, and the slowest 20% of trials on the psychomotor vigilance task all loaded onto the same latent AC factor, and this factor was strongly correlated with WMC and gF. Thus, AC abilities are reliably related to WMC and gF.

As noted above, a critical aspect of AC is the ability to ensure that goal and task-relevant information is actively maintained in PM in the presence of interference and distraction. Thus, within the overall working memory system, AC is needed to ensure that task-relevant items are being actively maintained and attentional capture from internal and external distractors is prevented. With any lapse of attention it is likely that attention will be captured by salient stimuli due to the task goal being displaced from PM and resulting in erratic and reduced performance.

In general, there are two main types of lapses of attention (internal and external) both of which can derail the current train of thought. One potent form of internal distraction is mind-wandering or daydreaming. It is generally quite difficult to sustain attention on a task for a length of time (especially if the task is boring). A great deal of prior research suggests that

---

*p* < 0.05 level. (B) Confirmatory factor analysis for WMC, gF, AC, and off-task thoughts. 
Ospan = operation span; Symspan = symmetry span; Rspan = reading span; 
Raven = Raven Advanced Progressive Matrices; LS = letter sets; Anti = antisaccade, 
SARTacc = accuracy in sustained attention to response task; SARTsd = standard deviation of reaction times in the sustained attention to response task; PVT = psychomotor vigilance task; AOff = off-task thoughts in antisaccade; SOff = off-task thoughts in the SART; POff = off-task thoughts in the PVT. All paths and loadings are significant at the 
*p* < 0.05 level.
participants report mind-wandering during many cognitive tasks and that the degree of mind-wandering varies as a function of task variables such as time on task, task complexity, and task difficulty (McVay & Kane, 2010; Smallwood & Schooler, 2006). Importantly, mind-wandering rates correlate with task performance such that performance is lower when participants report that they were mind-wandering on the preceding trial compared to when participants report that they are currently focused on the task (McVay & Kane, 2010; Smallwood & Schooler, 2006). A number of recent studies have shown that low WMC individuals mind-wander more than high WMC individuals, and this variation in mind-wandering partially mediates the relation between WMC and AC (e.g., McVay & Kane, 2009, 2012a, 2012b; Robison & Unsworth, 2015; Unsworth & McMillan, 2013, 2014a). For example, McVay and Kane (2009) found that low WMC individuals reported more mind-wandering during the SART than high WMC individuals, and importantly that mind-wandering rates partially mediated the relation between WMC and performance on the SART. Subsequent work by McVay and Kane (2012a) and Kane & McVay (2012) has found that mind-wandering rates across various tasks (Stroop, SART, reading comprehension) correlate quite well and load on the same latent factor, and this latent mind-wandering factor correlates well with latent WMC and AC factors and mind-wandering mediated the WMC-reading comprehension relation. In follow-up research we found that individual differences in mind-wandering were due to a combination of factors including WMC, interest in the current task, and motivation to do well on the task (Unsworth & McMillan, 2013). Importantly, we found that the WMC–mind-wandering relation was independent of interest and motivation suggesting that low WMC individuals’ deficits in AC and susceptibility to mind-wandering were not simply due to a lack of interest or motivation, but rather reflected a real cognitive deficit that arises on tasks requiring focused attention and working memory processes. Indeed, recent research has found that mind-wandering occurs during WMC (Mrazek et al., 2012; Unsworth & Robison, 2016) and gF (Mrazek et al., 2012; Unsworth & McMillan, 2014b) tasks and mind-wandering rates are negatively related with overall task performance.

Variation in mind-wandering and WMC has also been found in more ecological contexts examining everyday attentional failures. For example, Kane et al. (2007) had participants perform WMC tasks in the laboratory and then participants carried PDAs for a week. Periodically throughout the day the PDAs would beep and participants would have to answer a
variety of questions about whether they had just been mind-wandering. Consistent with laboratory assessments of mind-wandering, Kane et al. found that low WMC individuals experienced more mind-wandering in daily life when their current task required concentration, was challenging, or was effortful. Similarly Unsworth, Brewer, and Spillers (2012) had participants perform a number of tasks in the laboratory (WMC, AC, prospective memory, retrospective memory) and then carry a diary around for a week logging their various cognitive failures. We found that WMC and AC assessed in the laboratory predicted everyday attentional failures such that low WMC individuals reported more mind-wandering than high WMC individuals. In a subsequent analysis of the data focusing only specific types of attentional failures, we (Unsworth, McMillan, Brewer, & Spillers, 2012) found that most attention failures occurred either in the classroom or while studying. Like Kane et al. (2007), we found that WMC and AC predicted everyday attentional failures that seemed to require a high degree of focused and sustained attention, but did not predict all types of attentional failures. Thus, low WMC individuals found it more difficult than high WMC individuals to sustain their attention on challenging and demanding tasks leading to attention failures (i.e., more mind-wandering). However, on tasks that did not require a great deal of effort, WMC was unrelated to mind-wandering, suggesting boundary conditions under which AC processes are needed (see also Kane, Poole, Tuholski, & Engle, 2006).

In addition to mind-wandering, lapses of attention can also occur due to potent external distraction such as a loud banging, a honking horn, or a colleague playing their music too loud. Like mind-wandering, AC abilities are needed to protect and maintain task-relevant information in working memory against these potent distractors. Note here we are particularly talking about distraction that not only occurs in the environment, but is also irrelevant to the task at hand. To assess this we (Unsworth & McMillan, 2014a) had participants perform a number of WMC and AC tasks in the laboratory. During the AC tasks we periodically asked participants about their current attentional state. Similar to McVay and Kane (2012a) we asked if participants were thinking about the current task or mind-wandering. In addition we also asked if participants were distracted by information in the external environment (Stawarczyk, Majerus, Maj, Van der Linden, & D’Argembeau, 2011). The idea being that low WMC individuals will be more likely than high WMC individuals to have their attention captured by both internal distractors (mind-wandering) and potent external distractors (such as loud noises or flickering lights while trying to sustain their attention
on the task at hand. We found that mind-wandering and external distraction were correlated at the latent level ($r = 0.44$; see also Unsworth, McMillan, et al. (2012) for a similar demonstration in everyday attention failures) and both were correlated with WMC, AC, and gF. In fact, the shared variance among external distraction, mind-wandering, and performance on the attention control tasks was strongly correlated with WMC. Indeed, as shown in Fig. 6B, susceptibility to off-task thoughts (here a combination of external distraction and mind-wandering) is related to WMC, AC, and gF suggesting that low ability individuals are more likely to have their attention captured by internal and external distraction. In follow-up research we have found that the extent to which WMC is related to mind-wandering or external distraction is somewhat dependent on whether potent external distractors are present (Robison & Unsworth, 2015). Specifically, when participants perform a task in a quiet room with little distraction, WMC seems to be related to mind-wandering. However, if distraction is present (in the form of irrelevant auditory information), then WMC seems to be related to external distraction, rather than to mind-wandering. Thus, WMC prevents attentional capture to mind-wandering and external distraction in a context-specific manner.

Collectively these results suggest that AC abilities are needed to prevent attentional capture (to both internal and external distraction) and to protect important, yet fragile, information in working memory. Building on this line of reasoning, we have suggested that a key aspect of AC that relates to WMC is whether one can consistently apply control across trials. That is, trial-to-trial variability in AC is critically important. High WMC individuals are better able to consistently sustain attention on task than low WMC individuals, resulting in more fluctuations and lapses of attention for low WMC individuals than high WMC individuals. Evidence consistent with this notion comes from a number of recent studies which have shown that low WMC individuals have more slow reaction times (RTs) and more variability in RTs during AC tasks than high WMC individuals (McVay & Kane, 2012b; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Unsworth, Redick, et al., 2010; Unsworth et al., 2012c; Unsworth, 2015). For example, Unsworth (2015) found that variability of RTs in AC tasks (but not variability in RTs on lexical decision tasks) correlated with WMC and gF. Furthermore, variability in RTs (particularly slow RTs) on AC tasks predicted mind-wandering rates (both in and out of the laboratory), WMC, and gF. Thus, the consistency of AC may be the key factor that relates to WMC and other cognitive abilities. Indeed, recently
Adam, Mance, Fukuda, and Vogel (2015) found that low WMC individuals experienced more trial-to-trial fluctuations in performance on a visual working memory task than high WMC individuals, suggesting that inconsistency in AC is a likely reason for poorer performance seen by low WMC individuals on various working memory tasks.

If consistency (or inconsistency) of AC is a critical factor, then one natural question is what gives rise to fluctuations in AC? Recently we have suggested that individual differences in the functioning of the locus coeruleus norepinephrine system (LC-NE) may be a key reason for individual differences in WMC and AC (Unsworth & Robison, 2015b). Briefly, the LC is a brain stem neuromodulatory nucleus that is responsible for most of the NE released in the brain, and it has widespread projections throughout the neocortex including frontal areas (Berridge & Waterhouse, 2003; Samuels & Szabadi, 2008). The LC also receives major inputs from the prefrontal cortex (particularly the anterior cingulate cortex) suggesting a reciprocal connection between the LC-NE system and frontal cortex (Arnsten & Goldman-Rakic, 1984; Jodo, Chiang, & Aston-Jones, 1998; Rajkowski, Lu, Zhu, Cohen, & Aston-Jones, 2000). Given these wide projections throughout neocortex, the LC-NE system may be particularly important in modulating representations in frontal cortex based on attentional control demands (Aston-Jones & Cohen, 2005; Cohen, Aston-Jones, & Gilzenrat, 2004). A great deal of recent research suggests that there is an inverted-U relationship between LC tonic activity and performance on various cognitive tasks such that at intermediate levels of tonic LC activity attention is focused and performance is good. But at high or low levels of tonic LC activity, attention is unfocused and performance is worse. Accordingly, we (Unsworth & Robison, 2015b) have suggested that low WMC is related to a dysregulation of LC activity such that low WMC individuals demonstrate more fluctuations in tonic LC activity than high WMC individuals. To examine this we utilized pretrial baseline pupil diameter as an indirect index of tonic LC activity (Aston-Jones & Cohen, 2005; Eldar, Cohen, & Niv, 2013; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Rajkowski, Kubiak, & Aston-Jones, 1993) during a visual arrays change detection task (Unsworth & Robison, 2015a). As shown in Fig. 7A, we found that error trials especially for small set sizes (set sizes 1 and 2) were associated with lower pretrial baseline pupil diameters than correct trials, suggesting that prior to the occurrence of an error participants were in a lowered alertness/arousal state. Additionally, we found that individual differences in WMC were correlated with trial-to-trial fluctuations in
pretrial baseline pupil diameter ($r = -0.35$), suggesting that low WMC individuals experienced more fluctuations in pupil diameter (and presumably tonic LC levels) than high WMC individuals. Indeed, shown in Fig. 7B are pretrial baseline pupil diameters for a typical high and typical low working memory capacity (WMC) individual. Thus, fluctuations in arousal can determine capacity at any given time. When arousal is optimal, capacity
will be at its maximum, but when arousal is too high or too low, capacity will be reduced leading to reductions in performance (Kahneman, 1973). This suggests the possibility that individual differences in AC abilities are due to variation in LC-NE functioning which are linked to deficits in frontal cortex. That is, the putative frontal deficits seen in low WMC individuals (Kane & Engle, 2002) may be partially due to differences in LC-NE functioning.

In addition to active maintenance of task- and goal-relevant information, AC abilities are needed in a host of situations. For example, Kane and Engle (2003) have argued that in WMC differences also arise in conflict resolution where even if the task goal is maintained, low WMC individuals are less able at resolving the conflict that arises between the task goal and more habitual behaviors than high WMC individuals (see also Meier & Kane, 2013, 2015). Additionally, low WMC individuals may experience broader deficits in AC such as inabilities to configure attention to particular objects or spatial locations compared to high WMC individuals (Bleckley, Durso, Crutchfield, Engle, & Khanna, 2003; Bleckley, Foster, & Engle, 2015). Furthermore, low WMC individuals may have particular problems filtering out irrelevant information (Vogel et al., 2005) which may be unrelated to lapses of attention and mind-wandering. For example, in a recent study we found that both mind-wandering and filtering predicted WMC, but that mind-wandering and filtering were unrelated and accounted for separate sources of variance in WMC (Unsworth & Robison, 2016). Thus, fully delineating the different components of AC abilities will be an important topic for future research. For now it is clear that AC abilities are an important facet of individual differences in WMC.

4.3 Secondary Memory

Although active maintenance of task- and goal-relevant information in PM is a critical component of working memory, in some situations that information will be lost from PM and will have to be retrieved from SM. In particular, when attention is removed from those representations in PM (due to attentional capture from internal or external sources or new incoming information), the representations will be displaced from PM and will have to be retrieved from SM to ensure further processing. Thus, a critical aspect of working memory and an important reason for individual differences in WMC is the ability to retrieve and reactivate information that could not be actively maintained in PM. Similar to prior research, we suggest that the success of retrieval will depend on a number of control and monitoring
processes that occur during encoding, retrieval, and postretrieval (Atkinson & Shiffrin, 1968; Nelson & Narens, 1990; Raaijmakers & Shiffrin, 1980). Specifically, we have relied on a simple search model where it is assumed that there are both directed and random components to the overall search process (Shiffrin, 1970). The directed component refers to those strategic processes that are under the control of the individual. These control processes include setting up a retrieval plan, selecting and utilizing appropriate encoding strategies, selecting and generating appropriate cues to search memory with, as well as various monitoring strategies and decisions to continue searching or not. The random component refers to the probabilistic nature of the search process in which a subset of information is activated by the cues (ie, the search set), and representations are subsequently sampled and recovered from this subset (Raaijmakers & Shiffrin, 1980; Shiffrin, 1970). We have argued that individual differences in WMC primarily represent differences in the use of the various directed control processes that allow for controlled interactions between PM and SM, and it is these control processes that result in the relation between WMC and SM abilities.

Evidence for an association between WMC and SM abilities comes from a number of studies which have shown strong relations at both the task and latent levels. For example, low WMC individuals perform more poorly than high WMC individuals on free recall (Unsworth, 2007, 2009a), cued recall (Unsworth, 2009b), item recognition (Unsworth, 2010a; Unsworth & Brewer, 2009), and source recognition (Unsworth, 2010a; Unsworth & Brewer, 2009). These differences are especially pronounced on tests that require self-initiated processing (Unsworth, 2009c). Furthermore, several studies have suggested that WMC differences in SM abilities partially account for the shared variance between WMC and gF (eg, Mogle, Lovett, Stawski, & Sliwinski, 2008; Unsworth, 2010a; Unsworth, Brewer, & Spillers, 2009; Unsworth et al., 2014). Indeed, as a further demonstration of the robustness of the SM relation with WMC and gF, shown in Fig. 8 is a confirmatory factor analysis examining data from 578 participants in our laboratory. As can be seen, delayed free recall, picture source recognition, and paired associates all loaded onto the same latent SM factor, and this factor was strongly correlated with WMC and gF. Additionally, we have found that WMC predicts a number of different everyday memory failures including forgetting information on an exam or homework and forgetting login or ID information (Unsworth, McMillan, Brewer, & Spillers, 2013). Thus, it is clear that there is a strong and important relation between individual differences in WMC and remembering from SM.
One potential reason for WMC differences on measures of SM is differences in encoding strategies and encoding abilities. As noted previously, encoding strategies such as rote rehearsal and coding were considered fundamental control processes in Atkinson and Shiffrin (1968) model. As such, encoding strategies should be a primary determinant of variability in memory performance and a reason for the WMC—SM relation. A great deal of prior research has shown that effective encoding strategy use correlates strongly with overall memory performance (Richardson, 1998). Furthermore, research has shown that individual differences in encoding strategies partially account for individual differences on measures of WMC (e.g., Dunlosky & Kane, 2007; Turley-Ames & Whitfield, 2003). In terms of the WMC—SM relation, several recent studies suggest that at least part of the correlation between WMC and performance on SM measures is

Figure 8 Confirmatory factor analysis for working memory capacity (WMC), fluid intelligence (gF), and secondary memory (SM). Ospan = operation span; Sym-span = symmetry span; Rspan = reading span; Raven = Raven Advanced Progressive Matrices; NS = number series; Ang = verbal analogies; DFR = delayed free recall; Pic-Sour = picture source recognition; PA = paired associates. All paths and loadings are significant at the \( p < 0.05 \) level.
due to differences in encoding strategies (Bailey, Dunlosky, & Kane, 2008; Unsworth, 2016; Unsworth & Spillers, 2010b). For example, Bailey et al. (2008) found that measures of WMC correlated with reported strategy use such that high WMC individuals were more likely to report using more effective strategies (e.g., imagery and sentence generation) than low WMC individuals. Importantly, Bailey et al. found that individual differences in strategy use partially accounted for the relation between WMC and SM measures (see also Unsworth & Spillers, 2010b). More recently we examined individual differences in WMC and encoding strategies on several delayed free recall tasks at the latent level (Unsworth, 2016). We found that WMC correlated positively \(r = 0.32\) with reported use of effective strategies (i.e., interactive imagery, sentence generation, and grouping), but not \(r = 0.01\) with ineffective strategies (i.e., passive reading and simple repetition). Furthermore, WMC did not correlate with variation in study time allocation \(r = 0.02\), suggesting that some aspects of controlled encoding (effective strategy use), but not others (ineffective strategy use and study time allocation), were related to WMC. Indeed, as shown in Fig. 9, high and low WMC individuals do not seem to differ in the use of ineffective strategies, but there are large differences in the use of effective strategies. High WMC individuals are more likely and better able to use effective strategies than low WMC individuals. Importantly, this variation in effective strategy use partially mediated the relation between WMC and SM performance. Specifically, WMC and SM abilities were correlated \(r = 0.41\), but

![Figure 9](image_url)  
**Figure 9** Proportion of reported strategy use as a function of strategy type (ineffective vs. effective) and working memory capacity (WMC). Error bars reflect one standard error of the mean.
Once variation in encoding strategies was partialed out the correlation dropped substantially ($pr = 0.28$). Thus, individual differences in WMC are related to the ability to select and utilize effective encoding strategies which is an important determinant of performance on measures of SM.

Not only is WMC important for properly encoding information, but WMC is also needed at retrieval (Spillers & Unsworth, 2011; Unsworth, 2007; Unsworth, Brewer, & Spillers, 2013; Unsworth, Spillers, & Brewer, 2012a, 2012b). Much of our earlier research examining WMC differences in retrieval was concerned with the idea that high and low WMC individuals differ in the extent to which they can focus their search on the desired information in SM. Relying on search models of recall (Raaijmakers & Shiffrin, 1980; Shiffrin, 1970), we suggested that one of the main reasons high and low WMC individuals differ in recall performance is because low WMC individuals are unable to focus the search as well as high WMC individuals (due to poorer use of probes/cues), and thus low WMC individuals search through a larger set of items than high WMC individuals. That is, low WMC individuals have larger search sets than high WMC individuals due to the inclusion of more intrusions (both previous list and extra-list) resulting in more proactive interference for low WMC individuals than for high WMC individuals (Kane & Engle, 2000; Unsworth, 2010b). We have argued previously that low WMC individuals have larger search sets because they rely on noisier context cues than high WMC individuals and thus sample from a much broader temporal distribution than high WMC individuals (eg, Unsworth, 2007; Unsworth & Engle, 2007) and are worse at using temporal context as a cue (Spillers & Unsworth, 2011). The net effect of having larger search sets is that the probability of sampling a correct target item is lower overall. Furthermore, according to search models of this type, given larger search sets, low WMC individuals should recall items at a slower rate (leading to slower recall latencies) and should be more likely to output errors (intrusions) than high WMC individuals. A number of studies have found just this pattern of results (ie, lower correct recall performance, longer recall latencies, and greater frequency of intrusions for low WMC individuals than for high WMC individuals) in a number of free (eg, Unsworth, 2007, 2009b, 2016; Unsworth & Engle, 2007) and cued (Unsworth, 2009a; Unsworth, Brewer, Spillers, 2011; Unsworth, Spillers, Brewer, 2011) recall paradigms. Thus, there is ample evidence suggesting that WMC differences in recall are, at least partially, due to differences in search set size.

We have further argued that the reason that low WMC individuals have larger search sets than high WMC individuals is because low WMC
individuals are poorer at selecting and implementing effective retrieval strategies to self-generate appropriate retrieval cues (Unsworth et al., 2013; Unsworth et al., 2012a, 2012b). Theoretically, controlled search processes are reliant on intact frontally mediated control processes (Atkinson & Shiffrin, 1968; Burgess & Shallice, 1996). These control processes are especially important to select appropriate retrieval strategies, to generate appropriate contexts to search, to elaborate on cues needed for search, to verify the products of the search, and to adequately use the products of the search to better focus the retrieval specification (Koriat, Goldsmith, & Halamish, 2008; Raaijmakers & Shiffrin, 1980; Shiffrin, 1970). Thus, these control processes and individual differences in WMC should be of vital importance when one is attempting to strategically search SM, and these control processes should be especially important during retrieval strategy selection and cue-elaboration phases where one must self-generate different contexts to search. To examine these notions we had high and low WMC individuals perform various fluency tasks in which participants must generate members of a category for a specified amount of time (for example, naming as many animals as possible in 5 min). Prior research with these tasks has shown that WMC is strongly related to overall performance (Rosen & Engle, 1997; Unsworth, Brewer, et al., 2011; Unsworth, Spillers, et al., 2011; Unsworth et al., 2012). Importantly, recent research suggests this relation is partially due to differences in retrieval strategies that participants use to generate items (Schelble, Therriault, & Miller, 2012; Unsworth et al., 2013). For example, we have shown that high WMC individuals generate more items and more clusters of related items than low WMC individuals when asked to generate animals for 5 min or friends on Facebook for 8 min (Unsworth et al., 2012, 2013). Examining how participants initiated retrieval suggested that high and low WMC individuals initiated retrieval in a similar fashion. Furthermore, examining the nature of the items retrieved suggested that high and low WMC individuals tended to retrieve in a similar fashion in that high and low WMC individuals retrieved a similar proportion of items from each of the different categories. Finally, although high and low WMC individuals reported using very similar strategies overall, high WMC individuals tended to rely more on their knowledge base to engage in general-to-specific searches than low WMC individuals and low WMC individuals were more likely to engage in a random search in which items were passively retrieved than high WMC individuals (see also Schelble et al., 2012). Importantly, these differences in reported retrieval strategy use accounted for the relation between WMC and number of
animals retrieved and between WMC and the number of clusters retrieved (Unsworth et al., 2013). Thus, differences in the ability to use retrieval strategies to self-generate retrieval cues seem to be an important reason for the relation between WMC and retrieval from SM. The notion that high WMC individuals are better at self-generating retrieval cues was directly examined in a second experiment where we had high and low WMC individuals perform the fluency task in the presence or absence of retrieval cues (Unsworth et al., 2013). We found that when no cues were present, high WMC individuals outperformed low WMC individuals consistent with prior research. However, when retrieval cues were present and participants were required to use the retrieval cues, performance was boosted and high and low WMC individuals retrieved the same number of items (see also Unsworth et al., 2012a, 2012b). Thus, these results suggest that WMC differences in retrieval from SM are partially due to differences in strategic search failures whereby low WMC individuals are less able to select and use retrieval strategies to self-generate retrieval cues.

Final aspects of controlled search that seem related to WMC are postretrieval monitoring and editing processes. After an item has been retrieved from SM, individual differences in WMC are related to the ability to effectively monitor the products of the search process and edit out intrusions (Lilienthal, Rose, Tamez, Myerson, & Hale, 2015; Rose, 2013; Unsworth, 2009b; Unsworth & Brewer, 2010a, 2010b). A number of prior studies have shown that low WMC individuals make more intrusions than high WMC individuals because they are poorer at monitoring the products of retrieval and correctly recognizing and editing out errors due to deficits in source monitoring (Lilienthal et al., 2015; Rose, 2013; Unsworth, 2009b; Unsworth & Brewer, 2010a, 2010b). Thus, low WMC individuals are not only more likely to generate intrusions (due to the use of poorer retrieval cues), but they are also less able to use source monitoring processes to determine the correct source and to effectively prevent intrusions from being recalled.

Collectively prior research suggests an important relation between WMC and SM abilities. These SM abilities refer to the ability to successfully encode information into SM and to recover information that was recently displaced from PM or to bring relevant items into PM. In order for information to be retrieved from SM it is critically important that information was successfully encoded in the first place and that appropriate retrieval cues can be generated to access the desired information and monitor the products of retrieval. All of these SM abilities seem critical to the WMC-SM relation as
in evidenced by recent research which suggests that the combination of encoding strategies, search efficiency, and monitoring abilities mediate the relation between WMC and SM (Unsworth, 2016).

5. MEASUREMENT OF WORKING MEMORY CAPACITY

Although there are many putative measures of WMC, we (and others) have primarily relied on complex working memory span tasks, simple span tasks, and visual arrays change detection tasks. Here we briefly outline what we think occurs during these tasks and what facets of WMC are primarily tapped by these tasks. For example, shown in Fig. 10A is a schematic depiction of the processes that occur during a typical version of the operation span task (or other complex span tasks). First, participants are presented with a math problem which they solve. Next, a to-be-remembered (TBR) item (here a word) is presented. With the presentation of the first word attention is focused on aspects of the first item and it is maintained in PM and participants engage in strategic encoding of the words (Bailey et al., 2008; Dunlosky & Kane, 2007; Turley-Ames & Whitfield, 2003; Unsworth & Spillers, 2010b). Depending on individual differences in WMC and task demands, these encoding strategies could be as simple as repeating the words over and over, or using more effective encoding strategies such as interactive imagery or creating sentences out of the words. At the same time, information maintained in PM is bound to the current context (temporal contextual as well as environmental context) creating item-context bindings which along with strategic encoding factors will be used during retrieval (eg, Davelaar et al., 2005; Lehman & Malmberg, 2013). Following presentation of the first word, the next math problem is presented. With the presentation of the math problem, the first word is displaced from PM as attention is switched to the math problem (eg, Craik & Levy, 1976; Unsworth & Engle, 2007, 2008). During the presentation of the math problem if there is any free time following the successful solution of the math problem participants will attempt to covertly retrieve the first word presented (McCabe, 2008; Rose, Myerson, Roediger, & Hale, 2010). This covert retrieval process serves to bring the item back into PM (ie, it becomes part of the current focus of attention) thereby strengthening the item and updating the item-context bindings (Loaiza & McCabe, 2012). With the presentation of the next word, participants can include the new word along with any other words covertly retrieved into the existing encoding strategy. Bindings will also
be created between the new word and context (item-context bindings) and between the new word and any words that have been covertly retrieved (item-item bindings; Lehman & Malmberg, 2013; Raaijmakers & Shiffrin, 1980). The idea that items are covertly retrieved from SM back into PM during complex span tasks is consistent with recent research demonstrating that during the encoding phase of complex span tasks there is significant hippocampal activation (Faraco et al., 2011). This hippocampal activation likely

Figure 10 Schematic depiction of typical trials on (A) operation span, (B) simple word span, and (C) visual arrays change detection tasks.
reflects covert retrieval processes that bring items back into PM from SM as well as the creation of bindings between items and the current context.

The process of displacement, covert retrieval and updating, and combination of covertly retrieved and new words likely continues until the recall period. At this point any items that are maintained in PM (ie, items currently being attended to) because they have been covertly retrieved or because there is no distractor activity to displace them will be recalled first if free recall is required. Following the unloading of items from PM, participants then initiate a search of SM for the current TBR items based on the use of temporal-contextual cues. Similar to prior work (Atkinson & Shiffrin, 1971; Raaijmakers & Shiffrin, 1980), it is assumed that during search of SM, PM holds the retrieval cues or pointers needed to access items in SM. In the current episodic memory tasks, temporal-contextual cues are used to define search sets that encapsulate the TBR items. The more precise the temporal-contextual cues are, the smaller the overall search set will be leading to a higher-probably of recall, a reduction in the number of previous list intrusions, and a decrease in recall latency (eg, Unsworth & Engle, 2007).

Although the majority of items are likely recalled from SM in complex span tasks, it is crucial to point out that PM processes that occur during encoding (ie, encoding strategies, covert retrieval, item-context bindings) are critical for performance, and thus these tasks represent a combination of PM and SM processes. Furthermore, given that prior research has demonstrated that mind-wandering occurs during complex span tasks and is predictive of overall performance (Mrazek et al., 2012), AC abilities will also be critically important during complex span tasks.

Similar overall processes are thought to occur in the performance of simple span tasks. Like complex span tasks, in simple span tasks participants are presented with a series of TBR items (such as words), and after a variable number of items participants are asked to recall the items in the correct serial order. Shown in Fig. 10B is a schematic depiction of the processes that occur during a typical version of a simple span task (here word span). Similar to complex span, with the presentation of the first word attention is focused on aspects of the first item and it is maintained in PM and participants engage in strategic encoding of the words, and information maintained in PM is bound to the current context creating item-context bindings. Because there is no intervening activity to displace items from PM, items are either recalled from PM or from SM depending on the number of items and on the way items are displaced from PM. Once the capacity of PM is exceeded, some items will be displaced from PM. In some situations the items will be
covertly retrieved back into PM and the item-context bindings will be updated (McCabe, 2008). Other times, the item will not be covertly retrieved, but a retrieval attempt from SM will occur during recall. During recall, items are unloaded from PM and temporal-contextual search of SM is undertaken to retrieve items that could not be maintained in PM. Thus, the similarity between complex and simple spans is that items must be recalled both from PM and SM. The main difference is that the majority of items in complex spans are displaced from PM and must be retrieved from SM, whereas for simple spans many items can be recalled from PM. Similar to complex span tasks, AC abilities are needed to sustain attention on the task and prevent mind-wandering and trial-to-trial fluctuations in attention.

From this framework we can also consider what happens in a typical version of a visual arrays change detection task. Shown in Fig. 10C is a schematic depiction of the processes that occur during a typical version of a change detection task. Participants are briefly presented with an array of colored squares followed by a delay period and then the test array. The participant’s task is to indicate whether the circled item in the test array has changed its color from the memory array. With the presentation of the array, attention is focused on the items to maintain them in PM. During the brief presentation of the array, participants may utilize various encoding strategies such as maintain all of the items or just a subset (Bengson & Luck, 2016; Cusack et al., 2009) or rely on various perceptual grouping strategies (Peterson & Berryhill, 2013; Woodman, Vecera, & Luck, 2003). During this time, bindings of item to context and spatial location are created and maintained. Furthermore, depending on whether other irrelevant items are presented or if the number of items presented exceeds capacity, filtering operations may come into play to filter out the distracting items (Cusack et al., 2009; Vogel et al., 2005). If the number of items presented exceeds capacity, some target items will be displaced from PM, and if needed a search of SM will be needed to attempt to retrieve them. During the delay period, AC abilities are needed to actively maintain the items in PM and to prevent lapses of attention and mind-wandering (Adam et al., 2015; Unsworth & Robison, 2015, 2016). Upon presentation of the test array, items in PM are assessed. If the cued item is not in PM, then a search of SM ensues in an attempt to retrieved the target item. Although these tasks primarily reflect PM capacity and AC abilities (eg, Shipstead et al., 2014; Unsworth et al., 2014), SM abilities are also needed on occasion in these tasks. That is, prior research suggests that performance on these tasks is susceptible to proactive interference (Hartshorne, 2008; Shipstead & Engle, 2013), suggesting that
on some trials participants attempt to retrieve items from SM. If the item is not in PM, cannot be retrieved from SM, or if retrieval is not attempted, then participants will resort to guessing. Across trials, AC abilities are needed to prevent mind-wandering and trial-to-trial fluctuations in attention (Adam et al., 2015; Unsworth & Robison, 2015, 2016). Thus, these tasks primarily reflect a combination of PM capacity and AC abilities, with a smaller contribution coming from SM abilities.

Collectively, various working memory tasks rely on a combination of PM capacity, AC abilities, and SM abilities. These tasks differ in the extent to which they draw on these different facets of WMC resulting in differential relations among themselves and with other tasks. That is, we suggest that all immediate memory tasks measure the same basic set of processes, accounting for their predictive power across a wide range of tasks. Yet we acknowledge the tasks differ in the extent to which they draw on these different processes resulting in slightly different indices of individual differences in WMC.

6. HETEROGENEITY OF WORKING MEMORY CAPACITY LIMITATIONS

Throughout we have suggested that working memory is not a unitary system, but rather is composed of multiple distinct, yet interacting, facets and that each of these facets are important for higher-order cognition. Specifically, the current review suggests that PM capacity, AC, and SM abilities contribute to individual differences in WMC and are each part of the reason why WMC predicts high-order cognitive functioning so well. Collectively, prior research indicates the multifaceted nature of WMC and further suggests that rather than assuming that WMC limitations are the result of a single factor or process, we suggest that WMC limitations can arise for a number of reasons. Specifically, some individuals may have deficits in PM capacity which limits the number of items that can be distinctly maintained. Other individuals may have deficits in AC abilities resulting in lapses of attention (mind-wandering) and attentional capture whereby irrelevant distractors gain access to PM. Yet, other individuals may have deficits in SM abilities resulting in problems in encoding information into SM, retrieving information from SM, or correctly recognizing and editing out intrusions. Prior cluster analytic research supports these notions by demonstrating that some individuals have deficits in one process, but strengths in another, while still other individuals have deficits in all processes or strengths in all
(Unsworth, 2009a; Unsworth et al., 2014). These results provide important evidence that WMC limitations are multifaceted. The notion that individuals can be low or high in WMC for a number of reasons can potentially help resolve discrepancies in the literature where some studies find evidence for the importance of deficits in one facet (eg, PM), whereas other studies find evidence for the importance of another facet (eg, SM). These discrepancies could potentially be due to differences in the samples and/or working memory measures used where one facet is more represented than another leading differences in the resulting correlations. Future research should further examine the notion that WMC limitations and individual differences in WMC are multifaceted.

7. CONCLUSIONS

We have suggested that WMC and individual differences in WMC are multifaceted with differences arising due to variation in PM capacity, AC, and SM abilities. Although we have primarily focused on these three facets and recent research suggests these three facets mediate the relation between WMC and gF (Unsworth et al., 2014), we note that there are other important processes which individuals differ on and are likely important for WMC. These include integration, coordination, updating and attention switching, and binding operations (Bayliss, Jarrold, Gunn, & Baddeley, 2003; Halford, Cowan, & Andrews, 2007; Oberauer, 2002; Oberauer, Süß, Wilhelm, & Wittmann, 2003; Unsworth & Engle, 2008; Verhaeghen & Basak, 2005). Each of these processes has been linked to working memory and individual differences in WMC in prior research, and these processes have been suggested as possible reasons for the strong relation between WM and higher-order cognition. Future research is needed to determine the extent to which these processes (as well as other important processes) are related with PM capacity, AC, SM abilities, and overall variation in WMC. We suggest that there are multiple facets to working memory and to individual differences in WMC. To fully understand working memory processes and individual differences in WMC, we must strive to understand the operations of these different facets and how they interact.

REFERENCES


