The Ins and Outs of Working Memory:

Dynamic Processes Associated with Focus Switching and Search

Paul Verhaeghen, John Cerella

Syracuse University

Chandramallika Basak

Beckman Institute

Kara Bopp

Wofford College

Yanmin Zhang, William J. Hoyer

Syracuse University

Recent theories of working memory propose that working memory is not unitary, but subdivided into concentric regions that differ in the accessibility of the stored information. Cowan's (e.g., 1995, 2001) model has probably been most influential in this regard. This model proposes a hierarchical two-tier structure for working memory, distinguishing a zone of privileged and immediate access, labeled the focus of attention, from a larger activated portion of long-term memory in which items are stored in a readily available but not immediately accessible state. (In accordance with the terminology introduced by McElree, e.g., 2001, we define accessibility of an element in working memory by the time needed to retrieve it; availability is defined by the probability that the element is retrieved correctly.) The focus of attention is typically considered to be capacity-limited, and to contain a fixed number of items; the activated portion of long-term memory is not thought to be capacity-limited, but items stored in this structure are subject to interference and decay. Note that in this view, working memory is seen not as a separate cognitive system, but rather as an arena consisting of activated elements on which attentional processes operate; the 'central executive' (e.g., Baddeley, 1996) is no longer a structural element of the model, but it is equated with this set of processes.

One logical consequence of this proposed two-tier structure is the existence of a focus switch process (e.g., McElree, 2001; as far as we can determine, the term 'focus switch' has been coined by Voigt & Hagendorf, 2002). When the number of items to be retained in working memory is smaller than or equal to the capacity of the focus of attention, the item will by definition be immediately retrievable, and access times will be fast. When the number of items to be retained exceeds the capacity of the focus, however, the excess items will by necessity be stored outside the focus of attention and accessing the item for processing will necessitate a retrieval operation. This will slow down access time.

1. The magical number 2.5 plus-or-minus 1.5: Focus switching and the capacity of the focus of attention

Cowan's theory provides us with an operational definition of the capacity of the focus of attention, that is, its capacity is measured by the number of items that can be accessed immediately. The measurement can be made for any task that requires retrieval from working memory -- when the number of items to be retained is varied, a jump in response times will occur when the limit of the focus of attention is reached.

One interesting aspect of working memory revealed by focus-switching studies is that the capacity of the focus of attention is not fixed but adaptable, up to a point. Early research on working memory capacity has provided us with the magical number of seven items (plus or minus two; Miller, 1956). This estimate of the maximum capacity of working memory is probably too large. It was derived from forward digit span tasks, and it has been shown that these tasks are contaminated with the effects of rehearsal (Cowan, 2001). Cowan (for an extensive review, see Cowan, 2001) ascribes to the focus of attention a size of, not one, not seven, but four (plus or minus one). Evidence for his estimate is derived from a variety of experiments in a large number of research domains, including cluster sizes in free recall from episodic memory, the limits of perfect recall from immediate memory, proactive interference effects, the limits of cued partial reports, subitizing spans, multiple-object tracking, and the limits of consistently mapped search. The range of the evidence is staggering: Cowan's Table 1 lists 41 different 'selected key references' in 17 different research domains, all pointing to the magical number four.

Recent research using evoked potentials to directly measure the capacity of visual working memory as activation in the parietal cortex support this notion of a capacity limit of about four items (Vogel, McCollough, & Machizawa, in press; Vogel & Machizawa, 2004).

Earlier focus-switching studies, however, have converged on a smaller estimate. According to at least five prior studies, the focus of attention can accommodate only a single item at any given time (Garavan, 1998; McElree, 1998, 2001; Oberauer, 2002; Verhaeghen & Basak, 2005; a future downward estimate of that number is unlikely). As an illustration of this claim, **Figure 1** depicts the results from the latter study (Verhaeghen & Basak, 2005, Expt 1, college-age sample only). The subjects in this study were performing a version of the identity-judgment N-Back task (modeled after McElree, 2001; a schematic example of a single trial, consisting of 20 to-be-responded-to items, is represented in **Figure 2**). In this task, the subject sees a series of stimuli, presented one at a time on a computer screen; her task is to indicate whether the item currently presented is identical to the item presented *N* positions back. In our version, stimuli were shown on the screen in *N* virtual columns; these columns were additionally defined by stimulus color. Put more simply, the subject was instructed to indicate whether the item currently on the screen matched the item show previously in the same column. This columnized version of the task removes the control requirement of keeping track of the position of the stimulus in the stimulus series. The critical data are given by the response times (depicted in the bottom left panel of **Figure 1)**. It can be seen that the response time by *N* trace is close to a step function, with a fast time at $N = 1$, and then a jump to slower and statistically equal times for values of $N > 1$. Note that this jump, the focus-switch cost, is quite large – about 250 ms. These data, like similar data obtained by Garavan (1998), McElree (1998, 2001), and Oberauer (2002), strongly suggest that only one item is directly accessible at any given time, although the system clearly keeps more items in an available state, as exemplified in high accuracy for (at least) 4 items (see the bottom right panel **of Figure 1**).

We are thus faced with two conflicting accounts – there is a venerable tradition of research that points at a focus of size four, and a set of recent focus-switching studies that claim a focus of size one. How can these two types of data be reconciled? One possible explanation (Garavan, 1998) takes into account the nature of the tasks involved. In all the narrow-focus procedures, attention is directed serially to different elements that are either being encoded into working memory or retrieved from it. The serial requirement likely necessitates a controlled switch of attention to successive items; what distinguishes narrow-focus outcomes from wide-focus outcomes may then be the requirement to shift attention serially within either the stored representations or the to-be-encoded items in the narrow-focus contexts. Another possible explanation (Oberauer, 2002) is that the two outcomes derive from different memory structures. Oberauer argues for the existence of a concentric tripartite store, combining the architecture proposed by Cowan with the architecture proposed by McElree and Garavan. This architecture consists of (a) a focus of attention, containing a single chunk of information, namely the chunk actually selected as the object of the next cognitive operation; (b) a capacity-limited region of direct access corresponding to Cowan’s focus of attention, where a limited number of items is stored that are likely to be selected for subsequent processing; and (c) the activated part of long-term memory, that serves mainly to store information over brief periods of time.

The limitations imposed on the focus of attention by both the serial-attention explanation and the tripartite-architecture explanation are hard and fast. Both theories treat a focus-of-one as built into the working-memory machinery. In the light of attentional theory, we (Verhaeghen, Cerella, & Basak, 2004) considered the position that the focus of attention is limited to a single item excessively rigid, even under circumstances that require serial switching. The new working memory theories view working memory as an attentional system, and one of the basic suppositions of attentional theory is that attentional capacity can be allocated flexibly across the perceptual/cognitive field, following the needs of the participant, the demands of the task, or both (e.g., Kahneman, 1973; Wickens, 1984). We therefore considered a third possible resolution of the narrow-focus/wide-focus controversy: Perhaps the reported values represent the two ends of an underlying continuum of resource allocation. That is, perhaps the focus of attention can vary in size, shrinking to one item when all resources are channeled to the processing of a single item, and expanding up to a size of four when several items can be processed in parallel. On this view, a focus-of-one is not due to a structural limitation of the cognitive system, but rather to the way attention is distributed over the task and stimuli.

The simplest manipulation to test the immutability of the size of the focus of attention is a practice manipulation. This has not always been a successful strategy: Garavan (1998, Expt . 2) used an extended practice manipulation as a test for a resource-allocation explanation of the focus-of-one outcome in a symbol-counting task. In his study, a substantial focus-switch cost remained after practice; this result has since been replicated by Basak (2005). We took up Garavan's practice manipulation, and applied it to our columnized identity-judgment *N*-Back task, reasoning that this task may be simpler, especially because the channel or register to be accessed is predictable. Therefore we hoped that extended practice might lead to automaticity in some of the task components; this in turn should free up resources that could then be applied towards storage inside the focus of attention. The empirical test of this reasoning is straightforward: If the limit on the size of the focus of attention is structural, then we would expect that the *N*-back step function (with the step at $N = 2$) will remain intact over the course of practice (although the height of the step may be diminished). If, however, the location of the step shifts over the course of practice (perhaps settling finally at $N = 5$), this would suggest that a focus-of-one is due not to a structural limitation, but to a resource limitation.

Five adults practiced the *N*-Back task for ten hours each, spread over five consecutive days. **Figure 3** depicts the evolution over sessions of average response time and accuracy (Verhaeghen, Cerella, & Basak, 2004). The critical data are again the response time data (top panel). It can be seen that at the onset of practice, the data conform to the focus-of-one account of working memory: There is a sudden jump from $N = 1$ to $N = 2$, and a flat RT profile thereafter. After ten hours of practice with the task, however, the pattern of response times differed considerably from that of the first session. There is no longer evidence for privileged access in the $N = 1$ condition. Rather, the response time increases linearly over the range $N = 1$ to $N = 4$, with a shallow slope of 30 ms/*N*. The breakpoint is now situated at $N = 5$: The increment here is about twice as large as that extrapolated from the linear trend. In other words, it appears that with extended practice participants were able to expand the focus of attention to accommodate four items.

Our result then, suggests that with extended practice the focus of attention can be expanded to hold four items, even in a task requiring serial attention which initially limits processing to a single item. Two additional points are worth noting. First, our result strongly suggests that a size of four may be the ultimate limit of the size of the focus. This limit may be due to a limit on total activation in the relevant storage structures of the cortex (e.g., Vogel & Machizawa, 2004; see Usher, Haarmann, Cohen, & Horn, 2001, for a mathematical model of flexible allocation). Second, research has shown that not every task showing an initial focus of size one is amenable to a practice-related focus expansion (Garavan's symbol-counting task is not; Garavan, 1998; Basak, 2005, Expt 3); neither do all subject groups show this expansion (Basak, Cerella, & Verhaeghen, 2004, failed to obtain focus expansion in the *N*-Back task in a group of older adults).

Summarized, we maintain that the size of the focus of attention is not fixed, but that it may take any value between 1 and 4. (Put facetiously, the correct magical number would then be 2.5 plus-or-minus 1.5.) The exact size appears to be a function of yet-to-be-explored task characteristics, individual differences and their interaction.

2. A walk down working memory lane: Focus switching and retrieval dynamics within working memory

Our focus-switching process neatly slices working memory in half: a variable-size part characterizes by fast access times, and a surrounding part of unknown size with slower access times. This immediately raises the additional question what type of access processes operate within each of these parts. Interestingly, our studies reveal differential retrieval dynamics for the two zones.

2.1. Retrieval dynamics within the focus of attention

Our extended practice experiment allows us to examine retrieval processes within the expanded focus of attention. At first blush, this may appear to be a superfluous question: If all items within the focus are immediately accessible, one might expect that they are all accessed at the same speed. This is not necessarily true, however: In a limited-capacity system, items will compete for activation, and the competition might be fiercer when the system has to maintain a larger set of items. For instance, in support of his bound (four items plus or minus one), Cowan (1999) cites Trick and Pylyshyn (1994) on the subitizing span, and Fisher (1984) on parallel channels in consistently-mapped visual search; both are processes characterized by shallow, positive, load functions. Another memory process with this linear signature is short-term memory scanning (Sternberg, 1966). Mathematical models of this process are highly developed, and converge on the schema of parallel, self-terminating, limited-capacity memory access (Murdock, 1971; Ratcliff, 1978; Van Zandt & Townsend, 1993).

The available data suggest that access within the focus of attention confirms to Cowan's expectation: As stated above and as can be seen in Figure 3, there is a significant response-time-over-N slope of about 30 ms/item over the $N = 1$ to $N = 4$ range. The presence of a slope strongly suggests that the focus of attention is not content-addressable; rather, its contents need to be searched.

To investigate the possibility that this search process within the focus is of a parallel and limited-capacity nature, as might be expected from Cowan's theory, we conducted an ex-Gaussian decomposition of the response time distributions. Briefly, the ex-Gaussian model assumes that each response time can be represented as the sum of a Gaussian or normally distributed random variable and an independent exponentially distributed random variable. The ex-Gaussian distribution is described by three parameters: *mu* and *sigma* are the mean and standard deviation of the normal distribution and *tau* is the mean of the exponential distribution. *Mu* and *sigma* determine the location of the leading edge of the distribution; *tau* reflects slow responses at the tail of the distribution (the skew). The reason for examining this decomposition is that Hockley (1984, Expt. 1) has shown that the ex-Gaussian signature of the Sternberg search process is a linear increase in tau over memory load, while *mu* and *sigma* remain constant. **Figure 4** (middle panel) presents the results of the ex-Gaussian decomposition for Session 10 of the Verhaeghen, Cerella, and Basak study (2004). Clearly, the linear increase in response time over *N* in the expanded focus of attention is solely due to an increase in the skew of the distribution, as expected when a limited-capacity parallel search process is involved. The striking resemblance in response time by load slopes of these three operations ─ our version of *N*-Back, memory scanning, and subitizing ─ and the identical signature of the *N*-Back and memory-scanning tasks in ex-Gaussian space suggest that the processing involved in all three tasks may be traceable to the same modus operandi within the focus of attention. This is a hypothesis that merits more research.

We note that the conceptualization of within-focus search as a parallel limited-capacity process might further explain why the capacity of the focus is not fixed. If access occurs in parallel but capacity-demanding channels, the number of supportable channels is likely to be a function of the total resources available to the working memory system. If a task is very demanding, the residual may allow only a single active channel. As a task becomes automated, resources are freed up, allowing the system to open more channels to simultaneous search.

2.2. Retrieval dynamics in working memory outside the focus of attention

What about search outside the focus of attention? As we have presented it here so far, the story is remarkably consistent and simple: Both Verhaeghen and Basak (2005), and Verhaeghen, Cerella, and Basak (2004, Session 1) obtained statistically flat slopes over values of N larger than 1 (**See Figure 1 and Figure 3**). Therefore, the items outside the focus of attention appear to be directly content-addressable, much in the way elements stored in long-term memory are (see also McElree, 2001).

This result needs to be qualified, however. First, Verhaeghen and Basak (2005) observed a flat slope only for college-age adults; older adults showed a slope of about 70 ms/item. Second, Basak (2005, Expt 1) found a clear slope of about 240 ms/register in focus-switch costs over *N* in an adapted version of Garavan's counting task. Third, Basak (2005, Expt 4) found a clear slope of about 90 ms/*N* in a version of the N-Back task in which the item to be accessed was not the item in the *N*th back position, but could be any of the items in the position 1 to *N* back (making this a hybrid between an *N*-Back task and a memory scanning task). Taken together, these results suggest that although the region outside the focus of attention can be content-addressable, it will likely only be so under the best of circumstances – precisely predictable switching performed by a group of high-functioning subjects. Under more normal circumstances, a search process is initiated.

This search process is clearly slower, by a factor of at least 3, than the limited-capacity parallel search process that seems to govern the focus of attention. Additionally, the search process outside the focus has a different ex-Gaussian signature than the search process inside the focus, as can be seen in the right-hand panel of **Figure 4**. Whereas the search slope within the focus of attention is entirely due to an increase in the skew parameter, search outside the focus seems to involve the mean and standard deviation of the Gaussian parameters as well. Thus, the search process outside the focus is clearly not a capacity-limited parallel search. Rather, its signature resembles that of a visual search task (Hockley, 1984), which is typically considered to be serial and self-terminating. More research is needed before we can with affirm this conclusion with great confidence, but it is an intriguing hypothesis.

3. The executive suite: Focus switching and control processes in working memory.

The role of the focus-switching process in working memory seems to be an important one. Focus switching is obviously not the only executive control process operating on working memory. It seems necessary, then, to investigate the relationship that focus switching has with these other control processes. More specifically, if we want to make the claim that focus switching is a cognitive primitive, we should be able to delineate this process from other control processes.

What are these control processes? Empirical attempts to fractionalize central executive processes using the method of confirmatory factor analysis (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Engle, Tuholski, Laughlin, & Conway, 1999; Oberauer, Süβ, Schulze, Wilhelm, & Wittmann, 2000) point at the existence of families of processes. The most ambitious of these projects has been undertaken by Miyake and colleagues. Miyake et al. (2000) investigated the separability of three types of executive processes derived from the neuropsychogical/attention literature – resistance to interference (note that Miyake et al. label the latter process ‘inhibition of prepotent responses’; we prefer the term coined by Dempster; see Dempster & Corkill, 1999), information updating and monitoring, and task switching. Miyake's three-factor model indeed provided a good fit to the covariance matrix, indicating factor separability. At the same time, the intercorrelations between the latent factors were quite high, ranging between .42 and .63, indicating considerable commonality.

Before we examine some of our data pertaining to focus switching and these three other processes, we should discuss one potential artifact in our data, namely the confounding of focus switching with working memory load.

3.1. Focus switching and working memory load

One obvious draw-back of our favorite task, the identity-judgment N-Back task, is that the jump from $N = 1$ to $N = 2$ which we attribute to focus switching is confounded with an increase in working memory load. We note that in this task, the focus-switch cost is a step function, so that the cost, after the initial jump from $N = 1$ to $N = 2$ does not increase with a further increase in working memory load. This qualifies, but does not eliminate the criticism – there may be something special about increasing the load from its minimum to any other value.

We have, however, conducted two experiments that suggest that the focus-switch cost cannot be due entirely to the increase working memory. For the first of these, we turned to a new paradigm, namely a repetition-detection task (Bopp, 2003). In this task, participants are presented with a series of stimuli, numbers between 1 and 16, presented one at a time on a computer screen. Within each series, one of the stimuli is shown twice; the participant's task is to indicate the identity of the repeated stimulus. Progress through the series is self-paced, so that response times and accuracy can be measured. Bopp manipulated focus-switching by having more than one series (labeled 'channels'; Bopp implemented a 1-channel, 2-channel and 3-channel condition) on the screen, each in a different location and a different color; the subject has to find the repeat within each series. Working memory load was manipulated as the length of the series. In the increasing-load condition, akin to the load increase in the N-back task, each series was 5 numbers long. This implies that for the 1-channel condition, the total load is 5 numbers; for the 2-channel version, the total load is 10 numbers; and for the 3-channel version, the total load is 15 numbers. In the near-constant-load condition, we always showed a series of 16 or 15 digits, so that for the 1-channel condition, the series was 16 digits long; for the 2-channel condition, each series was 8 digits long; and for the 3-channel condition, each series was 5 numbers long. Results are illustrated in **Figure 5**. A focus-switch cost is present in both versions of the task. Importantly, the near-constant-load condition shows a sizeable focus-switch cost of about 200 ms, and then no further increase in response times. This clearly demonstrates that the focus-switch cost is not solely due to an increase in working memory load. It also demonstrates that increased load does likely play a role in the size of the effect under circumstances when it is confounded with focus switching; an attempt to create more purified measures might be in order.

In a second experiment, we (Basak, 2005, Expt 2) had participants perform a modified version of Garavan’s symbol-counting task. In this task, subjects keep a separate running count of different shapes that appeared in a random sequence on a computer screen. Garavan used two shapes and hence two counts; Basak used up to four different shapes and hence four counts. Presentation is self-paced, allowing for the recording of response times. Response times are typically about 500 ms slower when the stimulus shape is changed from the previous trial than when both successive stimuli have the same shape. These results are typically interpreted as showing that participants keep separate mental counters active for each stimulus shape; the cost to switch between counters in working memory is the focus-switching cost. In her version of the task, Basak gave subjects a random start-up count to update from (e.g., the count of circles would start with 8; the count of squares with 4, etc.; the next circle seen would then update the circle count to 9, the next square would update the square count to 4, etc.). To manipulate load, Basak introduced a condition in which subjects were additionally given a start-up count for a 'passive' shape that they were told to ignore for the rest of the duration of the trial (i.e., instances of these to-be-ignored shapes would pop up, but subjects were told not to update their count for this shape). At the end of the trial, they were required to give the updated count for each of the to-be-updated shapes and recall the start-up count for each of the to-be-ignored shapes. If working memory load determined the focus-switch cost, the cost should increase with the presence and the number of passive items to be retained. It did not. An even more remarkable result is shown in **Figure 6**: the time to switch the focus of attention from an active shape to another active shape does not increase when a passive shape intervenes; neither does the time to return to the same shape suffer from the intervening presence of a passive shape. This demonstrates that under certain circumstances, working memory load may have no influence of the focus-switching cost at all -- as if the additional items are totally inert (see also Oberauer, 2003).

This may also be the place to note that one other simple load-related artifactual interpretation of the focus-switch cost can be eliminated outright, namely a speed-accuracy trade-off to deal with the presence of additional items in working memory. Verhaeghen and Hoyer (in press) found a negative correlation between the focus-switch cost in response time and the cost in accuracy, indicating that individuals who slow down most are also more likely to have the largest decrease in accuracy, contrary to what would be expected from a speed-accuracy trade-off mechanism.

3.2. Focus switching and resistance to interference

Resistance to interference is a crucial control process for working memory and likely one of the major determinants of working memory capacity (e.g., Hasher & Zacks, 1988; Kane & Engle, 2002). In tasks in which the focus presumably holds only a single item, the focus switch cost occurs at the moment when interfering items are introduced, hence contaminating this cost with the need for resistance to interference. The results reported in the previous paragraph already speak to this issue – a focus switch cost still occurs when the total number of items in working memory remains constant.

Another intriguing result that suggests that interference is unlikely to be a main source of the focus-switching effect emerges from an examination of the pattern of auto-correlations between items in the Verhaeghen and Basak (2005) study. (This analysis was not published in the original paper; the analysis was Basak's idea). The assumption is that if the subject is able to resist interference well current performance would be attuned to and hence influenced only by the item in the N^{th} position back. Hence, in a time-series analysis of response times, the autocorrelation function should show a clear spike in the N^{th} back position, indicating that performance on the current item is related to performance on the item in the N^{th} back position, at the exclusion of all others. Autocorrelation values should be close to zero for the other positions (with the possible exception of multiples of N, where lingering activation may still be present). This is indeed the pattern that was found (see **Figure 7**): Only the correlation between the current item and the item (multiples of) *N* back was found to be significant. The working memory system then seems to be extremely efficient in retrieving only the one item it needs to focus on and suppressing all others.

3.3. Focus switching and updating

A second control process to consider in relation to focus switching is item updating. In all of the studies mentioned, after the basic processing (e.g., the comparison and identity judgment processes for our N-Back task) has been completed, the item currently residing in the correct ‘channel’ outside the focus of attention will need to be overwritten with the item currently residing in the focus of attention. If the focus-switch cost, or part of it, is attributable to the requirement to update the contents of working memory outside the focus of attention, then one might expect that the concatenation of the two processes would yield an underadditive interaction.

The identity-judgment *N*-Back task allows for an examination of the updating process in conjunction with the focus switch. That is, in our *N*-Back task, updating is only necessary when the item presented in the N^{th} position back is different from the item currently in focus (i.e., a ‘no’ answer); no updating is required when the two items are identical (i.e., a ‘yes’ answer). Therefore, at least part of the increase in response times from ‘yes’ to ‘no’ answers should be attributed to the updating process. In four experiments (Verhaeghen & Basak, 2005, Expt 2; Verhaeghen, Cerella, & Basak, 2004; Zhang & Verhaeghen, 2005, Expts 1 and 2), the need for updating did not interact with focus-switch costs; one experiment (Verhaeghen & Basak, 2005, Expt 1) yielded an overadditive interaction. The results strongly suggest that updating and focus switching are executed independently and in serial fashion, perhaps with the added cost of unlocking shared resources before updating can follow focus switching.

3.4. Focus switching and task switching

A third control process to compare with focus switching is task switching. Task switching obviously shares a switching requirement with focus switching. In focus switching, the task remains the same across trials, but items have to be swapped in and out of the focus of attention. In task switching, the task changes from trial to trial, but no storage (and therefore no swapping) of items is necessary. Switching between tasks, like focus switching, typically increases response times (e.g., Jersild, 1927; Rogers & Monsell, 1995). If the switching requirement drives (part of) the focus-switching effect, one would expect an underadditive interaction between task switching and focus switching.

We designed one study (Verhaeghen & Hoyer, in press; see also Verhaeghen & Basak, 2005, Expt 2) to explicitly both switching processes. Our paradigm of choice was a continuous calculation task. The task was modeled after a number-reduction task devised by Woltz, Bell, Kyllonen, and Gardner (1996). In the continuous calculation task, participants are presented with a string of single-digit numbers, only one of which is visible at any given time. The participants are instructed to perform a calculation on the first pair, type in their answer, combine this answer with the following item to perform the next calculation, and so on. Two types of calculations ('rules') were used. When the numbers differed by two, the participant reported the average of the two, and used that number to be combined with the next number shown on the screen ('midpoint rule'). When the numbers differed by one, the participant reported the next number in the up or down numerical sequence, and used that number to be combined with the next number shown on the screen (‘up-and-down rule’). To aid participants, all items for which the midpoint rule had to be used were presented in yellow on a black background; all up-and-down items were presented in blue on the same black background. Focus switching was manipulated by having the participant work on one continuous series (‘single’ condition, no focus switching) versus having the participant work on two series, each one shown in a different column on the screen (‘dual’ condition, focus switching). Task switching was manipulated orthogonally by either having the participant work according to a single rule throughout a trial (‘pure’ condition) or by mixing the two rules according to a predictable ABAB… schema (‘mixed’ condition). Additionally, we recruited a group of older adults for this experiment as well as a group of college-age adults, to test for age-related dissociations.

The results of our experiment are depicted in **Figure 8**. As can be seen, our subjects achieved perfect additivity of focus switching and task switching in response times; the aging manipulation was additive to both these manipulations. Thus, it seems that subjects combine task switching and focus switching independently and in a serial fashion. Note that focus switching seems to be the more effortful process – the focus-switch cost in this experiment (around 600 ms) was about three times as large as the task-switching cost. Additionally, the costs in response times for the two processes were correlated only slightly (.19), and this correlation did not reach significance, offering a further indication that the underlying processes must be largely independent.

The accuracy data are of interest here because they show an age-related dissociation: The drop in accuracy due to the focus-switch process is (much) larger in older adults than in younger adults, but the drop due to the task-switch process is not. This again suggests at least partial independence of the two processes.

A third piece of evidence for the independence of task switching and focus switching in this study was found in the parameters of the ex-Gaussian distribution. Focus switching produced a substantial increase in each of the parameters, indicating that both the leading edge and the dispersion (including the skew) of the distribution was affected. Task switching produced an increase in *mu*, or the leading edge of the distribution, only. Thus, task switching has the effect of shifting the distribution along the horizontal axis, without changing its dispersion, including the skew. This result can be explained by the insertion of a set of normally distributed processes that do not interfere with the computational requirements of the continuous calculation task itself. The results for focus switching suggest either that an additional ex-Gaussian process is added to the original distribution, or that several of the main component processes of the original distribution are slowed by a multiplicative factor.

4. The ins and outs of working memory: Conclusion

Working memory has be studied under a variety of aspects. Most researchers appear to be interested in static aspects, like the system’s capacity and structure. We presented results from a series of studies that are mainly concerned with dynamic aspects of working memory, namely the process that swaps items in and out of the focus of attention, retrieval dynamics, and the relation between focus switching and other executive control processes.

Summarized, our research supports the following conclusions. First, working memory contains at its core a zone of privileged access, the focus of attention. Second, depending on the task and on the allocation of resources (which is partially a function of experience with the task) this zone can hold between 1 and 4 items.. Third, items within this zone appear to be retrieved through a non-content-addressable retrieval mechanism, probably a parallel, self-terminating, limited-capacity search. Fourth, items stored in working memory beyond the focus of attention appear to be retrieved through a relatively slow, potentially serial search process, but can under ideal circumstances be subject to direct, content-addressable retrieval. Fifth, the focus-switching cost cannot be attributed completely to the increase in working load that often accompanies the switch. Sixth, the evidence to date suggests that the focus-switching process is largely independent from any of the other major executive control processes that are putatively operating in working memory – resistance to interference, updating, and task switching.

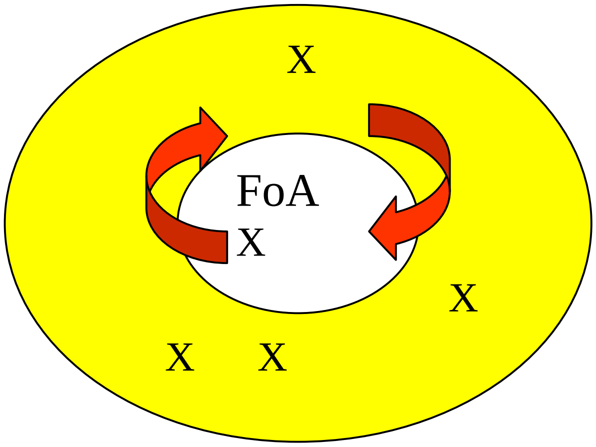
Our conclusions are still more tentative than we would like – the number of available experiments is relatively scarce. For instance, it is still unknown to what extent the present conclusions would generalize over a broad range of experimental paradigms. Another lacuna in our knowledge concerns the nature and determinants of the size of the focus-switch effect (Voigt & Hagendorf, 2002, provide first evidence for the influence of stimulus characteristics on focus switching; see also Zhang & Verhaeghen, 2005). Lastly, the role of focus-switching in standard working memory tasks, such as operation span, and in tasks involving multiple storage and processing demands, such as dual task coordination, needs to be elucidated.

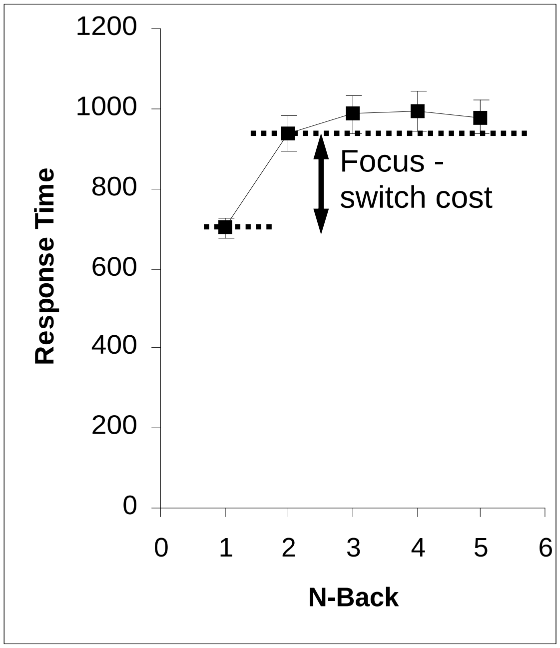
References

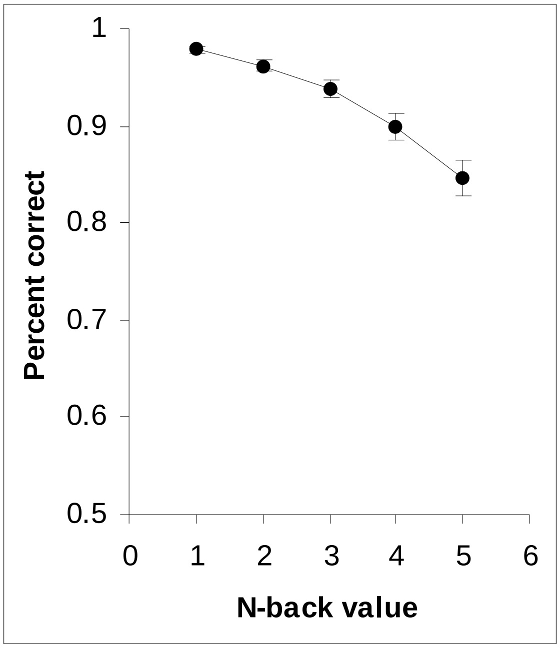
Acknowledgments

This research was supported in part by grants from the National Institute on Aging (AG-16201 and AG11451).

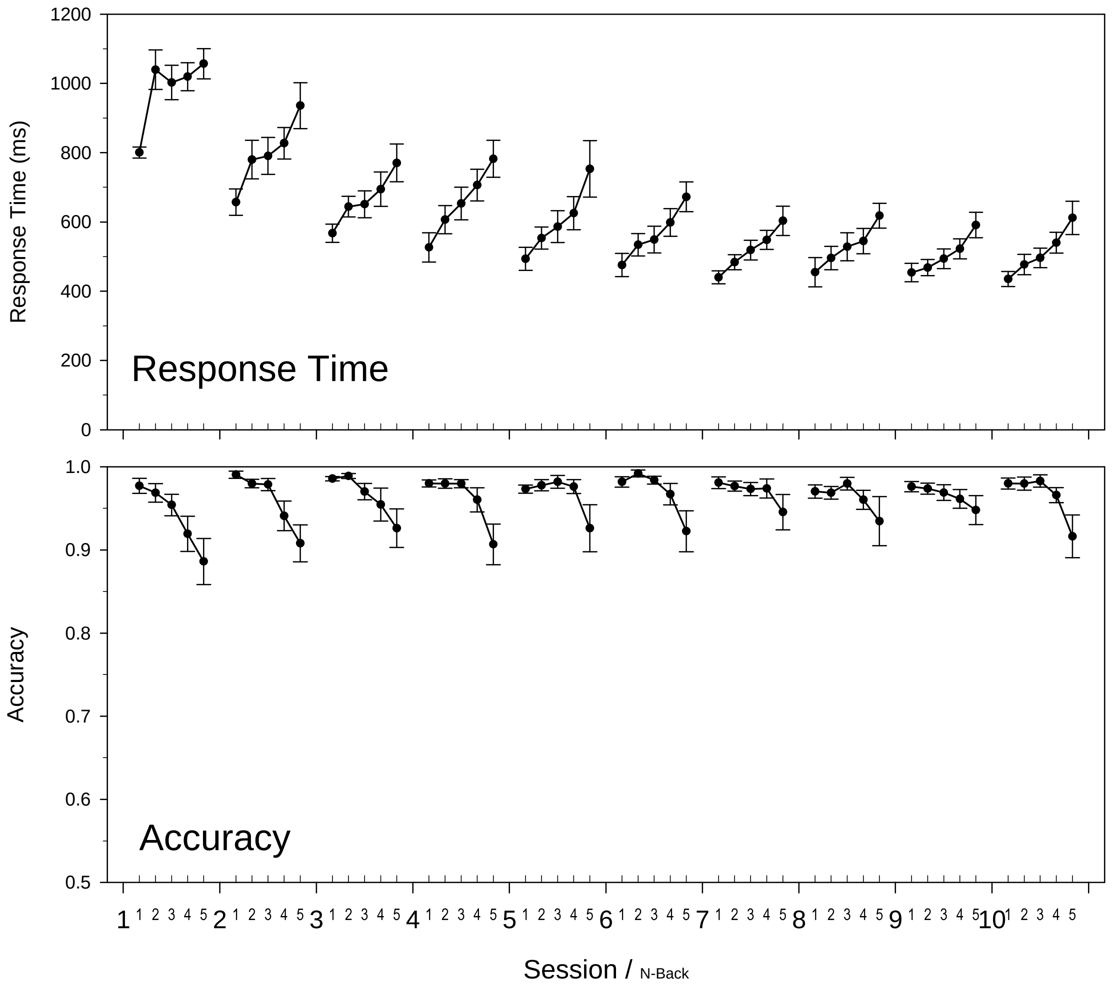
Figure captions

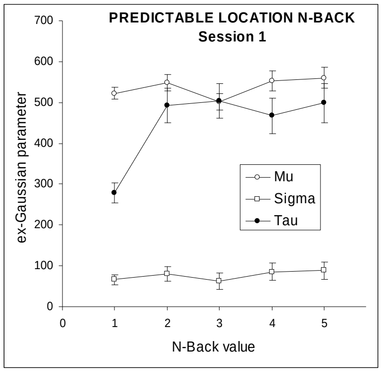

X
FoA
X
X
X
X

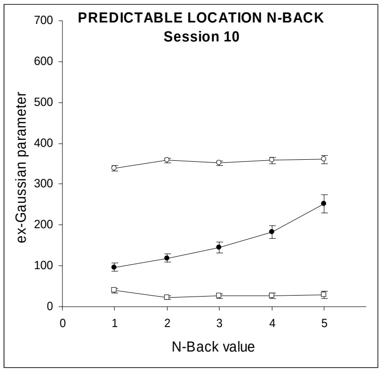

1200
1000
800
600
400
200
0
Response Time
0
1
2
3
4
5
6
N-Back
Focus -
switch cost

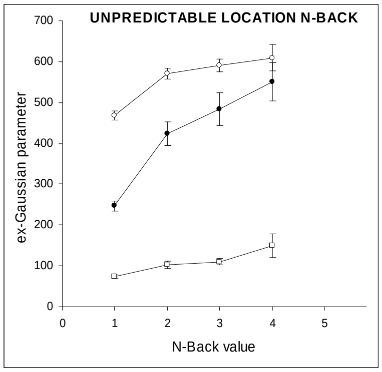

1
0.9
0.8
0.7
0.6
0.5
Percent correct
0
1
2
3
4
5
6
N-back value

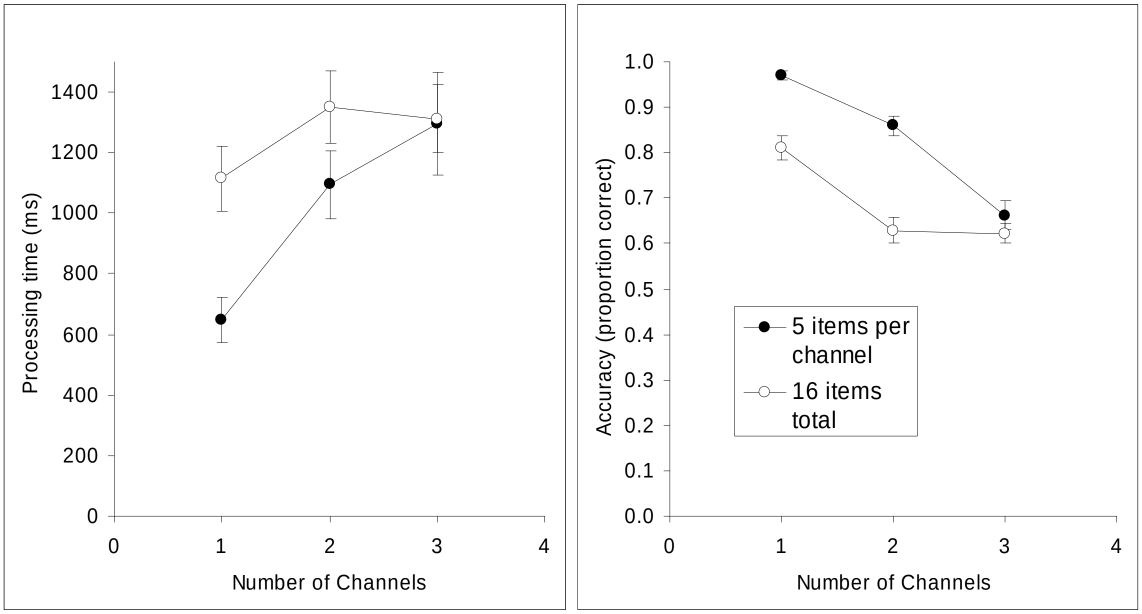
3	5	8	1
3	6	7	1
4	5	7	3
4	5	9	3
3	5	9	8
2	5	9	8

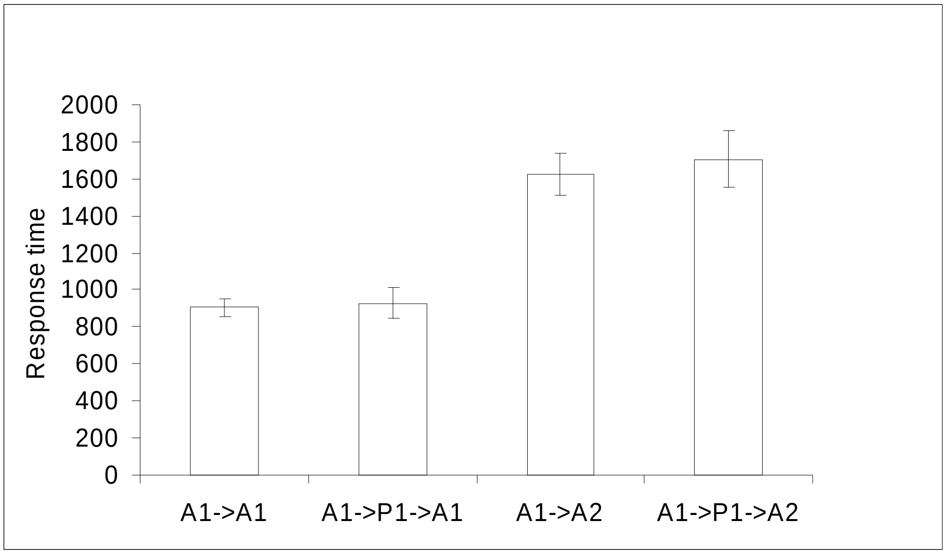

Response Time (ms)
1200
1000
800
600
400
200
0
Response Time
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
Accuracy
1 1 2 3 4 5 2 1 2 3 4 5 3 1 2 3 4 5 4 1 2 3 4 5 5 1 2 3 4 5 6 1 2 3 4 5 7 1 2 3 4 5 8 1 2 3 4 5 9 1 2 3 4 5 10 1 2 3 4 5
Session / N-Back

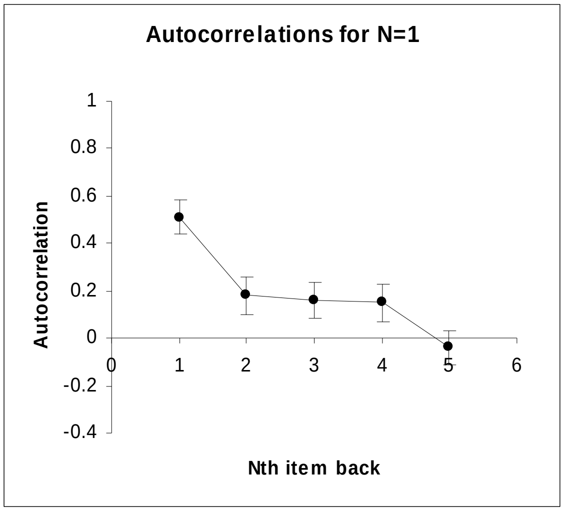

PREDICTABLE LOCATION N-BACK
Session 1
ex-Gaussian parameter
700
600
500
400
300
200
100
0
0
1
2
3
4
5
N-Back value
Mu
Sigma
Tau

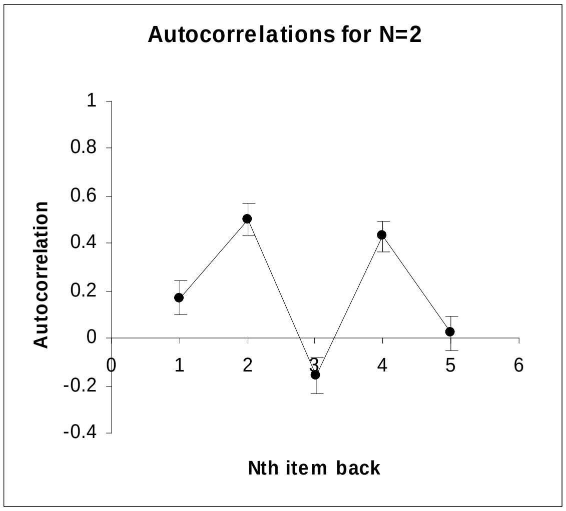

PREDICTABLE LOCATION N-BACK
Session 10
ex-Gaussian parameter
700
600
500
400
300
200
100
0
0
1
2
3
4
5
N-Back value

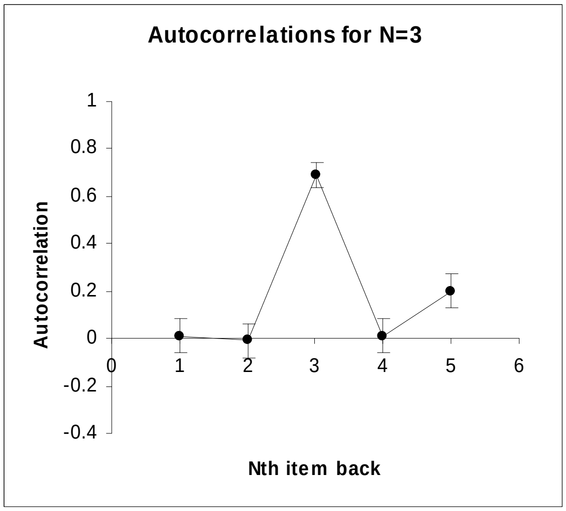

UNPREDICTABLE LOCATION N-BACK
ex-Gaussian parameter
700
600
500
400
300
200
100
0
0
1
2
3
4
5
N-Back value

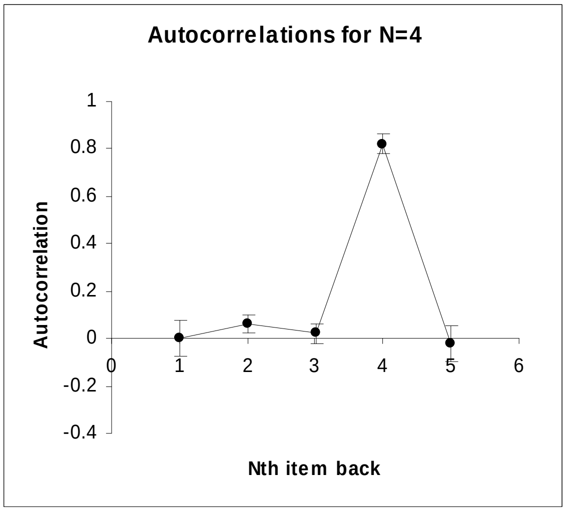

Processing time (ms)
0
200
400
600
800
1000
1200
1400
0
1
2
3
4
Number of Channels
Accuracy (proportion correct)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
1
2
3
4
Number of Channels
5 items per channel
16 items total

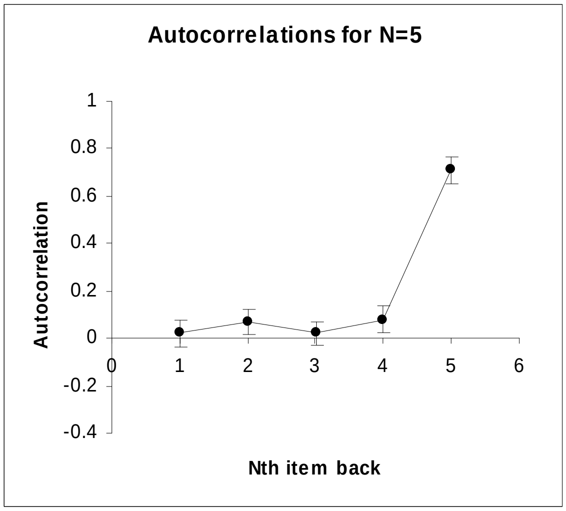

Response time
0
200
400
600
800
1000
1200
1400
1600
1800
2000
A1->A1
A1->P1->A1
A1->A2
A1->P1->A2

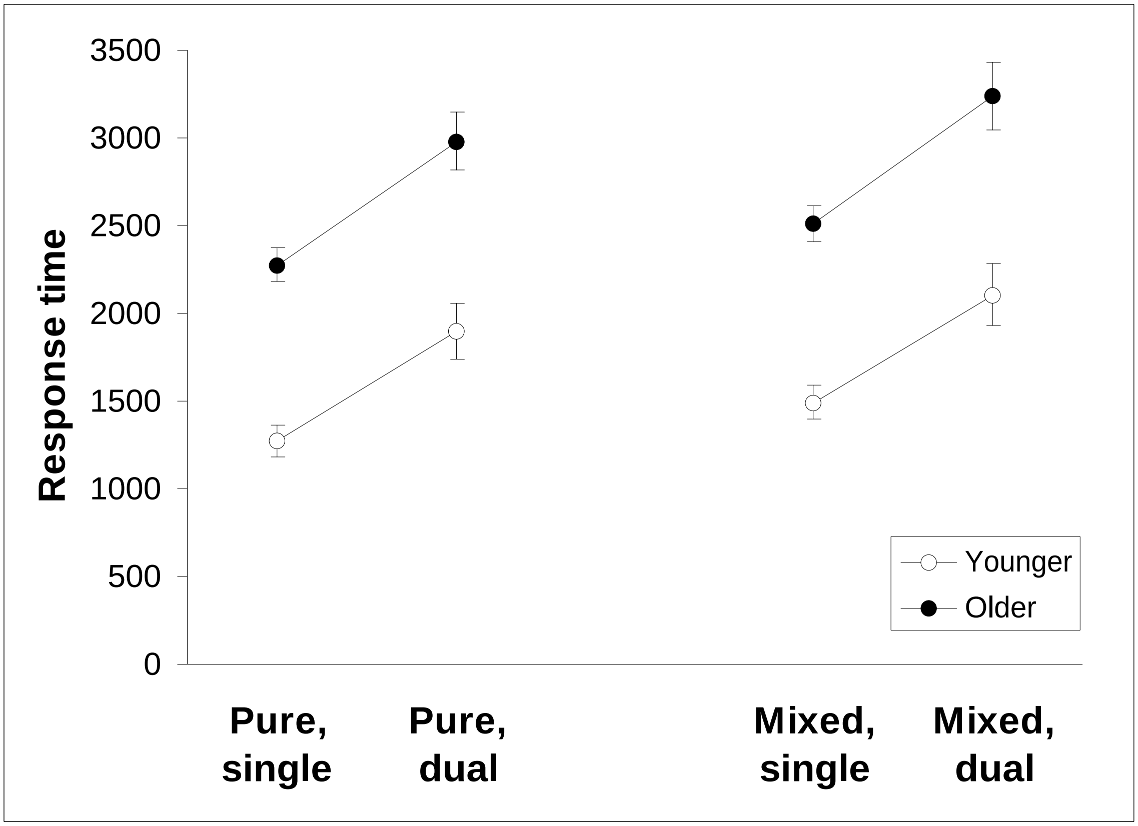

Autocorrelations for N=1
Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
1
2
3
4
5
6
Nth item back

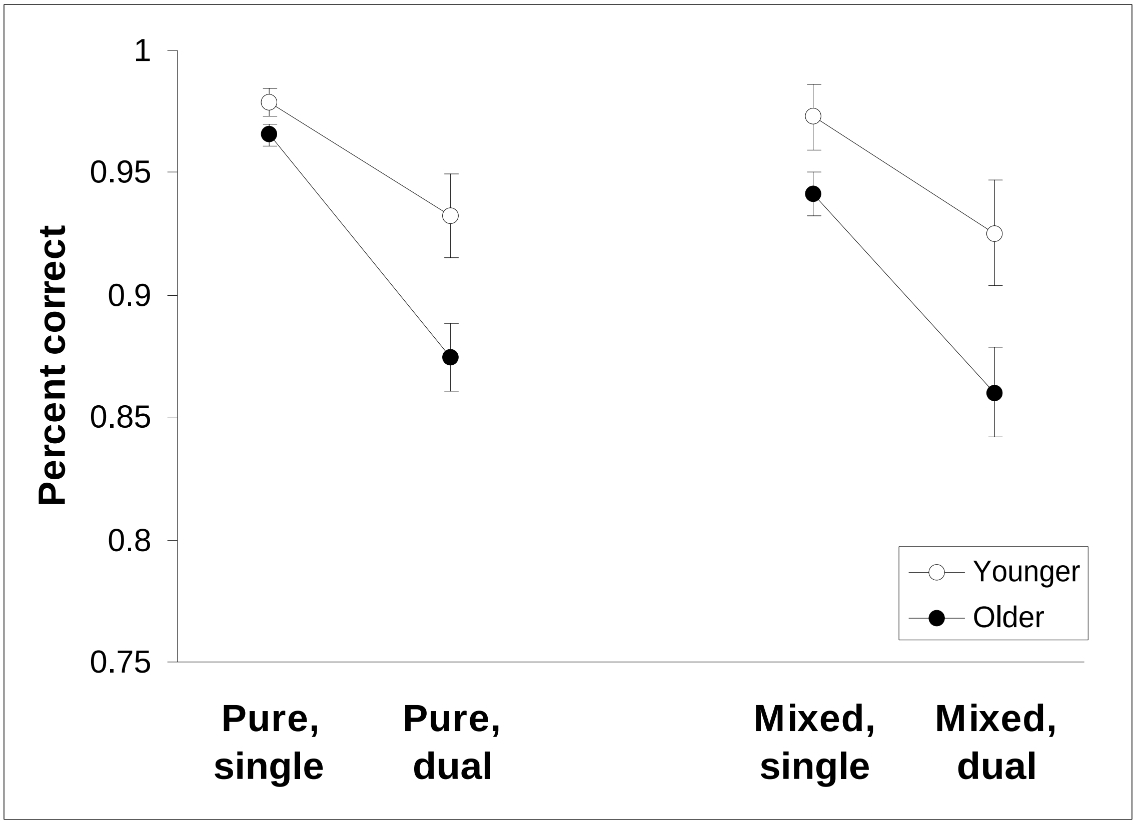

Autocorrelations for N=2
Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
1
2
3
4
5
6
Nth item back


Autocorrelations for N=3
Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
1
2
3
4
5
6
Nth item back


Autocorrelations for N=4
Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
1
2
3
4
5
6
Nth item back


Autocorrelations for N=5
Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
1
2
3
4
5
6
Nth item back


Response time
3500
3000
2500
2000
1500
1000
500
0
Younger
Older
Pure, single
Pure, dual
Mixed, single
Mixed, dual


Percent correct
1
0.95
0.9
0.85
0.8
0.75
Younger
Older
Pure, single
Pure, dual
Mixed, single
Mixed, dual