Atten Percept Psychophys
DOI 10.3758/s13414-016-1140-5

Visual working memory capacity increases between ages 3 and 8 years, controlling for gains in attention, perception, and executive control

Hrag Pailian[1,2] · Melissa E. Libertus[2,3] · Lisa Feigenson[2] · Justin Halberda[2]

Abstract Research in adults has aimed to characterize constraints on the capacity of Visual Working Memory (VWM), in part because of the system's broader impacts throughout cognition. However, less is known about how VWM develops in childhood. Existing work has reached conflicting conclusions as to whether VWM storage capacity increases after infancy, and if so, when and by how much. One challenge is that previous studies did not control for developmental changes in attention and executive processing, which also may undergo improvement. We investigated the development of VWM storage capacity in children from 3 to 8 years of age, and in adults, while controlling for developmental change in exogenous and endogenous attention and executive control. Our results reveal that, when controlling for improvements in these abilities, VWM storage capacity increases across development and approaches adult-like levels between ages 6 and 8 years. More generally, this work highlights the value of estimating working memory, attention, perception, and decision-making components together.

Keywords Visual working memory · Development · Attention and executive control

✉ Hrag Pailian
pailian@fas.harvard.edu

1 Department of Psychology, Harvard University, 33 Kirkland Street, 702 William James Hall, Cambridge, MA 02138, USA

2 Department of Psychological and Brain Sciences, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

3 Department of Psychology, Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260, USA

Our ability to make sense of a dynamic visual world requires the ability to form and store representations in Visual Working Memory (VWM). By maintaining the mental contents on which we perform computations even in the absence of visual input (Baddeley & Hitch, 1974; Flombaum & Scholl, 2006; Shinskey & Munakata, 2003), VWM functions as a gatekeeper between perception and complex cognition. As such, VWM significantly impacts various aspects of mental life (Vogel & Awh, 2008; Cowan et al. 2005).

In recent years, researchers have investigated the constraints that limit the storage of information in VWM. To study these limits in adults, researchers have relied heavily on the One-Shot change detection paradigm (Luck & Vogel, 1997; Phillips 1974), in which participants see an array containing varying numbers of objects displayed briefly (e.g., 100 ms). After an intervening blank display, participants see either a test array that is identical to the memory array or an array in which one object's features have changed (e.g., an object has changed from red to blue). The logic is that if the observer can maintain representations of all of the objects and their features in VWM, they will notice if a change has occurred. By varying the number of objects in the array and measuring response accuracy, the One-Shot task can return an estimate of the number of items, or amount of information, each individual can maintain in VWM.

Results from this One-Shot task suggest that adults are roughly equally accurate at detecting featural changes in displays containing from one to four objects (although the details of this point remain an active area of debate, see Brady et al. 2011). With displays containing more than four objects, performance usually drops precipitously (Luck & Vogel, 1997; see also Pashler, 1988; Phillips, 1974; Scholl & Xu, 2001; Sperling, 1960; Vogel et al. 2001). This decrease in performance is typically taken as evidence that the observer's VWM storage capacity limit has been exceeded. Although this

Published online: 25 May 2016

characterizes the overall group performance among adults, stable individual differences in capacity also are seen. These individual differences in VWM storage capacity have been found to correlate with measures of intelligence (Fukuda et al., 2010), attentional capture (Anderson, Laurent, & Yantis, 2011; Fukuda & Vogel, 2011), and the ability to filter irrelevant information during visual processing (Vogel, McCollough, & Machizawa, 2005).

VWM storage capacity and development

Given the importance of VWM for various aspects of cognition, characterizing the development of VWM is valuable for understanding age- and experience-related change in a wide range of cognitive abilities. However, attempts to assess developmental improvements in VWM have yielded conflicting results.

In one study, VWM was measured in infants using a visual preference task (Ross-Sheehy, Oakes, & Luck, 2003). Infants between the ages of 4 and 13 months sat before two screens, each of which presented a looping sequence of arrays containing colored squares. The two arrays were synchronized such that an array of, e.g., three squares appeared on each screen simultaneously for 500 ms, disappeared for 250 ms, and then reappeared for 500 ms, with this cycle repeating continuously. On one of the screens, the colors of all of the squares remained constant across flashes. On the other, one of the squares changed color on each flash. Ross-Sheehy and colleagues found that 4- and 6.5-month-old infants looked longer at the changing than the unchanging array when each array contained only one square. This suggests that these infants successfully remembered the color of the square over the 250-ms blank interval, compared it to the color of the square in the subsequent array, and detected a discrepancy between the observed and remembered arrays. In contrast, when the arrays each contained two or three squares, 6.5-month-olds failed to prefer the changing array, suggesting that they were unable to store the color features of two or three objects in VWM concurrently. By 10 months, infants preferred the changing array when each array contained one, two, three, or four squares, but not six. The authors suggested that VWM storage capacity is restricted to a single item between 4 and 6.5 months of age but undergoes rapid developmental change and reaches an adult-like limit of four items by about 10 months.

In older children, VWM has been studied using tasks more similar to those used with adults, including the One-Shot change detection paradigm. In contrast to the findings with infants, One-Shot tasks suggest a much more protracted developmental improvement in VWM storage capacity. For example, Cowan and colleagues (2005) presented elementary school-aged children with displays much like those of Luck and Vogel (1997). Children saw an array of colored squares for 250 ms, a blank 900-ms interval, and then a second array with a black frame now surrounding one of the squares. Children reported whether this probed square had changed color. This task yielded VWM storage capacity estimates of approximately 3.5 items in 7- to 8-year-old children, and of 4 items in 9- to 10-year-olds. In another study, Riggs and colleagues (2006) modified the task by increasing the encoding interval of the initial display to 500 ms and eliminating the frame cue. Their method returned VWM storage capacity estimates of 1.5 items in 5-year-olds, 2.9 items in 7-year-olds, and 3.8 items in 10-year-olds. Finally, Simmering (2012) presented 3-, 4-, 5-, and 7-year-old children with a One-Shot change detection task in which the encoding interval was increased to 2,000 ms for 3- and 4-year-olds (and remained at 500 ms for the two older age groups, as in Riggs et al., 2006). Simmering found that VWM storage capacity was 2 to 3 items in children aged 3 to 5 years, and roughly 4 items at age 7 years.

Although this work reflects recognition of the importance of studying the development of memory capacity, no consensus on this development has emerged. Whereas studies with infants suggest that VWM storage capacity reaches adult-like levels by around 10 months, studies of older children suggest that capacity continues to develop across the early school years. Further, absolute estimates of capacity have varied for early school-aged children. Simmering's estimate of VWM storage capacity in 5-year-olds was approximately twice that of Riggs et al. (2006), and Simmering found adult-like capacity several years earlier than Cowan et al. (2005). A further concern is that these studies have estimated VWM capacity in isolation from other abilities that may be undergoing development during these same ages and might affect performance in VWM tasks (e.g., attention, perception, decision making).

Against this backdrop, we first consider methodological differences in the developmental research and discuss their possible contributions. Next, we describe a novel method for studying VWM storage capacity in the early school years, and present data from this method while also including tasks that control for ancillary factors that might have contaminated previous VWM estimates.

Methodological differences in previous studies

Does VWM storage capacity asymptote in infancy, or continue to change during early childhood? Riggs et al. (2006) suggested that the capacity estimates obtained by Ross-Sheehy and colleagues (2003) may not reflect VWM and might rather reflect the minimal demands of the visual preference task used with infants. Infants' preference for changing arrays of three or four items might not require memory for all of the items, but instead draw on the successful storage of a subset or even just a single item, which—if combined across trials, would yield a significant preference for the changing array. However, Oakes, Messenger, Ross-Sheehy, and Luck (2009) found that when 6-month-old infants (who in earlier work exhibited successful change detection with a maximum of one item) were shown 3-item-displays in which all three items changed color on every iteration, infants still did not reliably look at the changing screen (for related findings, see Feigenson & Carey, 2005). This suggests that memory for a subset of an array is not sufficient to drive infants' visual preference in this task.

Riggs and colleagues (2006) offered another possible source for the discrepancy between findings with infants and older children—that visual preference tasks may recruit a more passive, less explicit form of memory than the One-Shot task. This more passive memory might suffice to drive visual preference for featural changes in 3- or 4-item arrays, but might not empower children's explicit identification of a changing target in the One-Shot task. Although this suggestion might help to account for the divergence in results obtained with infants (using the visual preference method) versus older children (using the One-Shot method), inconsistencies still remain across studies using various versions of the One-Shot task in older children. Some of these differences likely stem from variations in the parameters used in the One-Shot task, including stimulus timing, whether stimulus colors were chosen with or without replacement, number of items presented, memory probes using the whole display or just a single item, the formula used to quantify capacity, and whether capacity estimates were produced by averaging performance across set sizes or by taking the highest observed capacity estimate across set sizes (Awh, Barton, & Vogel, 2007; Cowan, 2001; Cowan, et al., 2006; Pashler, 1988; Rouder et al., 2008; Simmering, 2012; Simmering & Perone, 2013; Vogel, Woodman, & Luck, 2006). Reinforcing the idea that the One-Shot task may be strongly affected by such differences, we have found in our own work with adults that that the One-Shot paradigm produces highly variable capacity estimates even within a single participant. One-Shot VWM capacity estimates varied significantly within each individual observer depending on the parameters used, and large numbers of trials were required to return reliable estimates (Pailian & Halberda, 2015). Such differences might help to account for the differences in VWM storage capacity observed by Cowan and colleagues (2005), Riggs and colleagues (2006), and Simmering (2012), especially considering that our work in adults (Pailian & Halberda, 2015) suggests that the small number of trials presented to each participant (as is almost unavoidable when testing children) may have led to widely varying capacity estimates.

In addition to these methodological variations in the One-Shot task, an important issue when considering developmental change in VWM is that other abilities may be improving concurrently. If these other abilities are required by the tasks used to measure VWM, then observed improvements in capacity may actually reflect changes in these other abilities rather than in memory storage itself (Simmering, 2012).

One broad class of non-storage related abilities required by all VWM tasks is the ability to control visual attention. Because attention is needed to segment items from a visual scene, encode their features, and bind features to items (Treisman & Gelade, 1980; Kahneman, Treisman, & Gibbs, 1992; Chun & Turk-Browne 2007; but see Johnson et al., 2008), developmental changes in visual attention might have contributed to the developmental improvements observed in VWM capacity.

Similarly, changes in global visual processing may have influenced previous estimates of children's VWM capacity. For example, although the brief presentation of items in the One-Shot task is intended to minimize the verbal recoding of information and to disrupt any engagement of complex strategies, this short viewing time may, by preventing participants from attending each item individually, lead participants to recruit mechanisms of global scene processing (e.g., ensemble statistics) (Brady & Alvarez, 2015; Pailian & Halberda, 2015).

Finally, changes in executive control—the ability to maintain task focus and ignore irrelevant information—may influence children's VWM performance. Although the ability to avoid attending to irrelevant items (Cowan et al., 2010) and to rehearse displays covertly (Cowan et al., 2011) cannot entirely explain age-related differences in VWM storage capacity, these are only a few of the many executive processes involved in the One-Shot task. Other executive components, such as loading and purging items from VWM, switching attention between items, comparing remembered items to perceived items, and reaching a criterion for executing a response, may all contribute to the developmental changes in performance.

The present study

The present study had two goals. First, because our previous work with adults shows that the One-Shot task can return unstable estimates of VWM storage capacity, we aimed to develop a more stable measure of children's VWM storage capacity. Second, because it is important to account for developmental change in nonstorage-related abilities when measuring children's VWM, we estimated capacity while controlling for several key perceptual and cognitive abilities.

To meet these goals, we developed a new "Flicker" task (Halberda, Pailian, Wetherhold, & Simons, 2006; Pailian & Halberda, 2015) that uses displays like those in the One-Shot task (e.g., to-be-remembered array of items, followed by a probe array in which one item has changed). Although the One-Shot task allows only a single viewing of the to-be-remembered array, the Flicker task presents the two arrays in continuous alternation, whereas the observer searches through these alternations until they find the changing target (a changing target is present on every trial). In the Flicker task, by measuring an observer's reaction time to find the changing target across trials and by manipulating set size to estimate search rate, we obtain an estimate of VWM storage capacity. In our previous work with adults (Pailian & Halberda, 2015), we found that Flicker estimates of VWM storage capacity are highly correlated with those produced by the One-Shot task ($r = 0.70$, $p < 0.001$). However, exploration of the psychometric properties of these two tasks revealed that the Flicker task produced more reliable and stable estimates of VWM storage capacity in adults than the One-Shot task and that performance in the Flicker task was less sensitive to changes in task parameters (Pailian & Halberda, 2015).

The Flicker method has several potential advantages for studying VWM storage capacity in children. First, because every trial has a changing target, the task is intuitive and inherently rewarding. In contrast, the One-Shot task is harder for young children to understand (Simmering, 2012) and is less motivating, in part because there are many trials in which no change has occurred. Second, the Flicker task may provide some methodological benefits for unifying across the developmental literature as it combines features of the visual preference task (i.e., used with infants and involving a continuous stream of flickering presentations), and the One-Shot task (i.e., used with older children and adults and requiring an explicit response).

With these potential advantages in mind, we used the Flicker task to ask whether VWM storage capacity changes during childhood while also controlling for developmental changes in visual attention, global visual processing, and executive control. These latter abilities were measured using two additional tasks: Feature Search and Conjunction Search. This approach of examining patterns of performance across several tasks has not been taken in previous research on the development of VWM.

We indexed the exogenous control of attention using a Feature Search task in which children were asked to locate a color singleton among a set of identical distractors. Attention in this paradigm is primarily exogenously controlled, as the properties of the visual display cause the target to "pop out" and involuntarily capture attention (Treisman & Gelade, 1980). Although there are diverse views as to which paradigms truly measure exogenous attention (Barbot, Landy, & Carrasco, 2012), the Feature Search paradigm that we used is generally regarded to provide a suitable metric, because the selection of information appears to be stimulus-driven and automatic (Theeuwes, 1993). This type of automatic selection typically leads to "efficient" visual search, as reflected by nearly flat slopes for response times across distractor set sizes (Egeth, Jonides & Wall, 1972; Johnston & Pashler, 1990; Treisman & Gelade, 1980). Furthermore, this task is thought to recruit global visual processing, as attention need not be directed to each item individually.

We used a Conjunction Search task to index children's ability to endogenously control attention. Children again searched for a cartoon character in a field of distractor characters, but this time the target shared color and shape properties with the distractors (thus, our "Conjunction Search" task also might be considered a difficult shape and color combination search, as opposed to a pure conjunction of two simple features). Though much debate has surrounded the question of whether target identification in the Conjunction search task results from serial (Treisman & Gelade, 1980) or parallel processing of information (Wolfe, Cave, & Franzel, 1989; Wolfe, 1994), it is generally agreed that endogenous attention is required in Conjunction search (see Quinlan, 2003 for review). Because search tasks can be thought of as yielding a more-or-less continuous range from efficient search (e.g., feature search) to inefficient search (e.g., conjunction search), our two search tasks might be thought of as one more efficient search task and one less efficient search task. Critically, attention in the Conjunction Search paradigm is typically understood to be under voluntary control, directed by the goals of the observer, and relatively more focused (Wolfe, Cave, & Franzel, 1989; Wolfe, 1994; Duncan & Humphreys, 1989). Such control typically leads to relatively "inefficient" search, with response time slopes increasing as a function of the number of distractors present (Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989).

In addition to measuring the exogenous and endogenous control of attention, these Feature and Conjunction search tasks allowed us to estimate developmental changes in executive abilities. These included non-perceptual decision-making (estimated by the intercept of search slopes) and global visual processing (estimated by the mean of Feature Search response times). Our analysis approach was to control for these abilities when testing for developmental improvements in VWM capacity. After controlling for developmental changes in these factors (each of which could contribute to performance in the Flicker or One-Shot tasks), any remaining improvement in Flicker task performance would provide robust evidence that VWM storage capacity improves during early childhood. This approach is particularly important, given that changes in these abilities across development could create the false appearance of changes in VWM storage capacity across development. As a final developmental question, we compared children's performance to that of adults, in order to ask whether VWM storage capacity continues to improve significantly beyond early childhood.

Method

Participants

Seventy-two children participated: twelve 3-year-olds (mean [M] = 3.72 years, standard deviation [SD] = 0.17; 6 females), twelve 4-year-olds (M = 4.61 years, SD = 0.32; 6 females), twelve 5-year-olds (M = 5.51 years, SD = 0.23; 3 females), twelve 6-year-olds (M = 6.56 years, SD = 0.26; 6 females), twelve 7-year-olds (M = 7.45 years, SD = 0.3; 6 females), and twelve 8-year-olds (M = 8.21 years, SD = 0.2; 5 females). Twenty-one additional children participated but were not included in the final analyses due to refusal to participate in all three tasks (n = 5), parental or sibling interference during the tasks (n = 7), inattentiveness (n = 4), failure to understand task instructions (n = 4), or experimenter error (n = 1). Children received a small gift (e.g., book, t-shirt) to thank them for their participation.

Twelve adults also participated (M = 19.93 years, SD = 1.16; 8 females). One additional adult participated but was not included in the final analysis due to a visual impairment. Adult participants and parents of child participants gave written consent as approved by the university Institutional Review Board, and children provided verbal assent.

Apparatus

All tasks were programmed using MatLab Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and were presented on a 13.3” MacBook Laptop computer (288-mm × 180-mm). Viewing distance was unconstrained but averaged approximately 50 cm. On average, stimuli subtended 3 × 3 degrees of visual angle each and were presented against a solid white background.

Design and procedure

Each participant completed three tasks: Flicker Change Detection, Feature Search, and Conjunction Search; task order was counterbalanced.

Flicker change detection task

At the beginning of the Flicker task, participants were introduced to the Sesame Street cartoon character, Abby Cadabby. The experimenter explained that Abby was a fairy with the magical power to change the identity of other characters. The experimenter then told the participants that Abby would show them her magical power. As Abby raised her magic wand, a Sesame Street character appeared at the center of the monitor, enclosed in a circular black frame whose radius subtended 9.77 degrees of visual angle (from a viewing distance of 50 cm). This character remained visible for approximately 3 seconds. At this point, the experimenter pressed a button on the laptop keyboard, prompting Abby to lower her wand and the character to disappear. After 500 ms, Abby raised her wand again, and a different Sesame Street character appeared in the same location inside the frame. This identity-change sequence repeated until all 12 Sesame Street characters to be used in the task had been introduced (Elmo, Telly, Baby Bear, Cookie Monster, Big Bird, Snuffleupagus, Prairie Dawn, Zoe, Bert, Oscar the Grouch, Ernie, and Super Grover). These characters were chosen, because they were easily distinguishable on the basis of color and overall shape. On average, the characters subtended 3 × 3 degrees of visual angle; Abby measured 7.72 × 7.72 degrees from a viewing distance of 50 cm.

Next, the experimenter told participants that they would play a game in which they would see different numbers of characters on the screen. The experimenter explained that Abby would keep changing one of the characters into a different character, and then back again. For example, Abby might transform Zoe into Ernie, then back into Zoe, and so on, while all of the other characters remained constant. Participants were instructed to find and point to the changing character as quickly as possible.

For each practice and test trial, a set of characters was presented inside the circular black frame, with Abby outside the frame throughout (Fig. 1). On every trial, the characters inside the frame appeared for 900 ms. As Abby lowered her magic wand, the characters disappeared for 150 ms—long enough to prevent the use of iconic memory (Rensink, O’Regan, & Clark, 1997). Then, Abby lifted her wand and a second display appeared for 900 ms, with one character replaced by a character not previously in the array. The locations of the characters remained unchanged from the first display to the second. This looping, with consistent spatial positions, further minimized any contribution of iconic memory, as the continuous presentation of the displays was expected to overwrite the fragile iconic representations (Becker, Pashler, & Anstis, 2000). The sequence looped continuously until participants pointed to the changing character, at which point the experimenter pushed the space bar. Pressing the space bar caused the display to freeze; this allowed the participant to continue indicating the target character while the experimenter used the mouse to click on the character that the participant had pointed to. The computer recorded whether or not the target had been correctly identified by the mouse click. Errors were rare, as participants seemed motivated to continue searching until they had found the changing target, and all changes were very salient once detected.

All participants first completed six practice trials. During these, the set size increased by one additional character each time participants correctly identified the changing target. If participants erroneously pointed to a non-target character during these practice trials, the experimenter did not press the space bar to stop the alternating screens, but instead told them

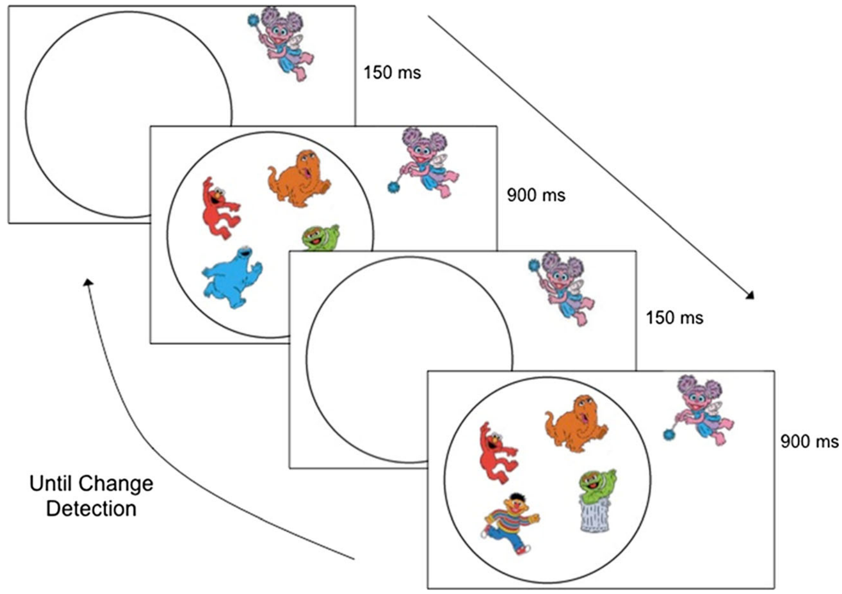


Fig. 1 Schematic of a Flicker change detection trial

that the identified character was not the target and that they should continue searching (e.g., "Let's keep looking; it looks like Abby is changing someone else. Can you find who is changing?").

Next, participants completed 24 test trials. These were similar to the practice trials, except that participants were presented with set sizes of four and six characters only (excluding Abby), randomly intermixed. The selection of characters for each trial and their spatial location within the circular frame were random across trials. No feedback was provided during the test trials, and participants were allowed to make only one pointing response.

Visual search tasks

Participants also completed two visual search tasks: a Feature Search task and a Conjunction Search task. These were modified versions of tasks used by Gerhardstein and Rovee-Collier (2002) to test 1- to 3-year-old children. In both search tasks, participants were introduced to novel cartoon characters that did not appear in the Flicker task. Participants were told that one of them (the target) would be presented amongst different numbers of distractors. Participants were instructed to search for the target and point to it as quickly as possible. The experimenter pressed the space bar when participants pointed and the amount of time from stimulus onset to the space bar press was recorded by the computer. For each participant, the target stimulus (either a red or a green character, counterbalanced across participants) was identical for the Feature Search and Conjunction Search tasks. Stimulus locations were randomly selected, and the number of distractors presented on each trial (2, 4, 8, or 12) was randomized.

In the Feature Search task, the target was presented among distractors that shared the target's shape but were of a different color (Fig. 2a). In the Conjunction Search task, the target was presented among two types of distractors. Some distractors shared the target's shape but had a different color, while others shared the target's color but had a different shape (Fig. 2a). Our goal was to create a difficult serial search task to measure the rate of controlled visual search (Fig. 3).

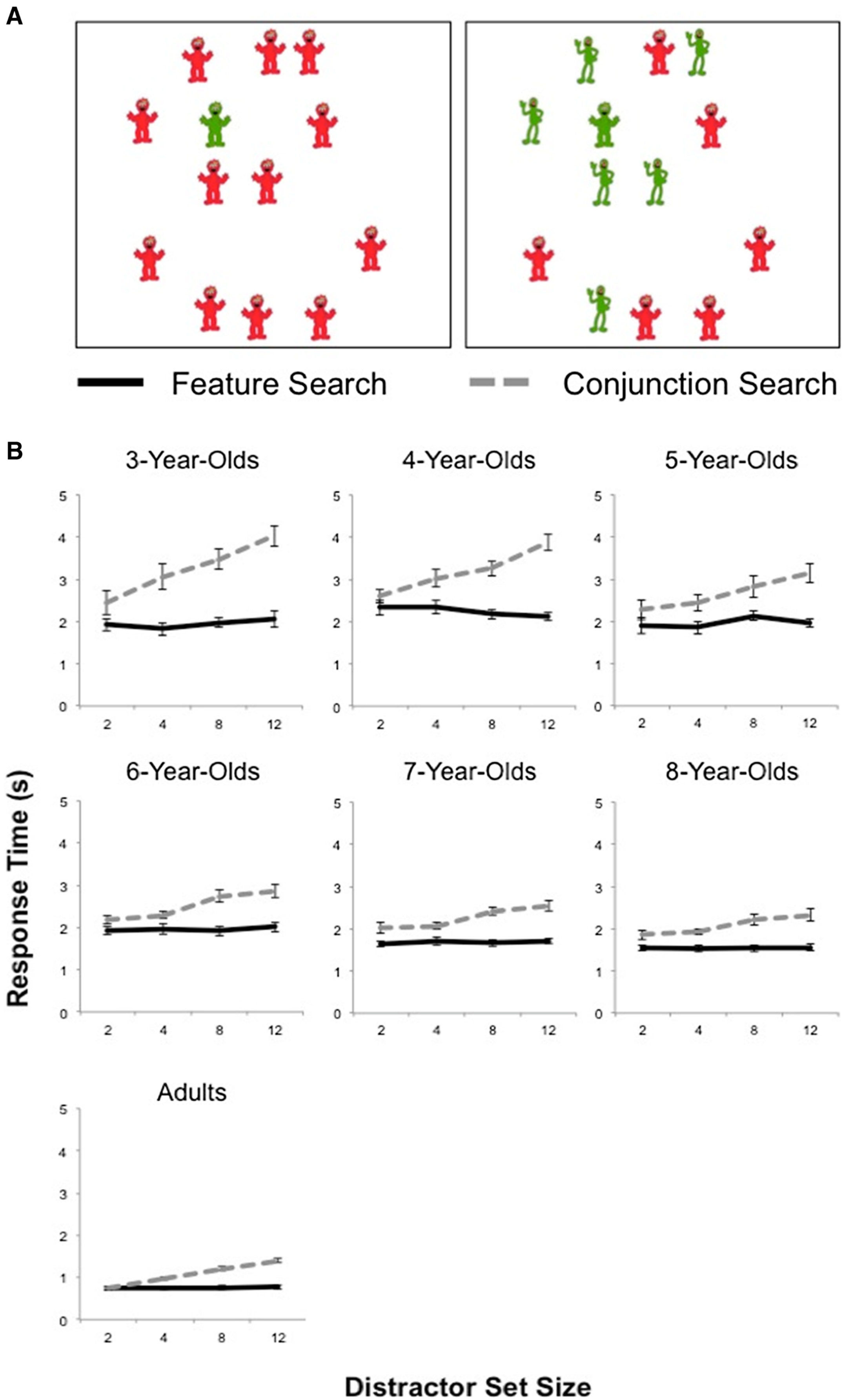
Before each search task, participants completed several practice trials. The practice trials were identical to their respective test trials, except that the number of distractors (2, 4, 6, 8) increased incrementally. On the first practice trial, the target was always presented among two distractors. If participants successfully found the target, the number of distractors then increased to four. If participants erroneously pointed to a distractor, the experimenter pointed to the target, and practice trials for that set size were repeated with different character locations until the target had been successfully located. On average, participants completed 4.99 practice trials of Feature Search (min = 4; max = 20) and 4.93 practice trials of Conjunction Search (min = 4; max = 7).

After the practice trials, participants completed 32 test trials in the Feature Search task and 32 test trials in the Conjunction Search task (8 each of set sizes 2, 4, 8, and 12 in random order). No feedback was given during test trials.

Data analysis

For our primary task of interest, the Flicker task, we focused our analyses on participants' average response time (RT). In addition, we present an exploratory analysis in which we transformed these RTs into an estimate of the item-based

Fig. 2 **A**) Schematics of arrays used in Feature Search and Conjunction Search tasks. **B**) Response times observed in both tasks decreased as a function of age ▶


A
Feature Search
Conjunction Search
B
3-Year-Olds
4-Year-Olds
5-Year-Olds
6-Year-Olds
7-Year-Olds
8-Year-Olds
Adults
Response Time (s)
Distractor Set Size
0
1
2
3
4
5
2
4
8
12

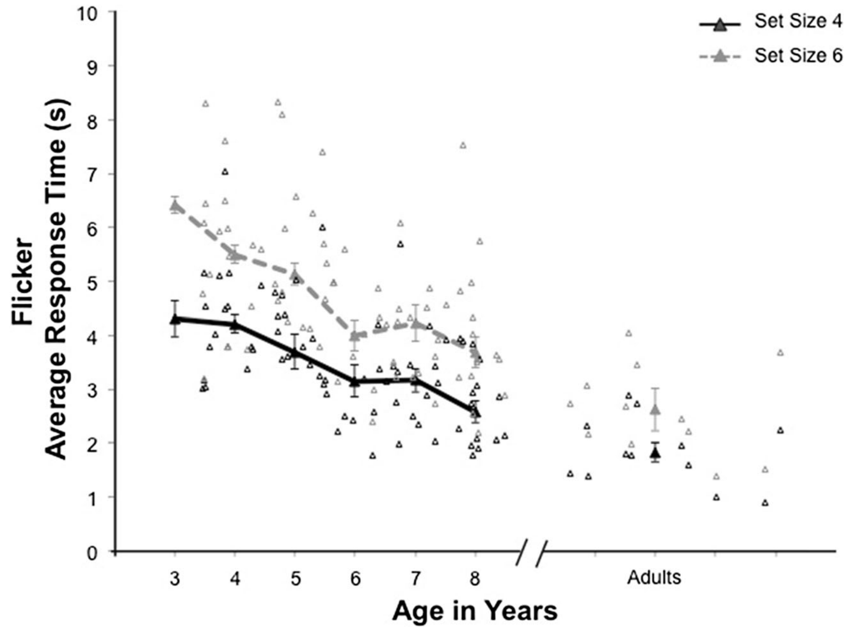


Fig. 3 Response times (s) observed in a Flicker task for set sizes 4 and 6 across all age groups. Filled symbols plot group means and error bars show the standard error for each mean value. Unfilled symbols plot response times for individual participants (Insert callout for Fig. 3 in the text—in numerical order)

storage limit of VWM (K), using an approach that we have used for adult participants (Pailian & Halberda, 2015).

For the Feature and Conjunction Search tasks, for each participant we calculated the slope and intercept of a linear regression line of response time (RT) across set sizes, as well as RT averaged across set size. These serve as measures of search efficiency (Treisman & Gelade, 1980; Wolfe, 1998). The intercepts of the search slopes for the Feature Search tasks estimate the fixed costs inherent to the task, including initiating the search process, pre-attentively processing shape and color, nonperceptual decision-making, and executing a motor response (Wolfe et al., 2002). Critically, these same abilities also are required by the Flicker task. For the Feature Search task, these processes typically constitute the majority of the response time; because the target differs from the distractors on the basis of an early visual feature (color), we expected that search would be highly efficient, as revealed by constant RTs regardless of the number of distractors present (Treisman & Gelade, 1980, Donderi & Zelnicker, 1969; Egeth, Jonides, & Wall, 1972).

Response time in the Conjunction Search task was expected to reflect the same processes as the Feature Search task (search initiation, pre-attentive processing of shape and color, nonperceptual decision-making, and motor response initiation), as well as some additional abilities. In the Conjunction Search task the target cannot be differentiated from the distractors on the basis of a single visual feature. As such, search is thought to require the deployment of focal attention in a goal-oriented manner, with participants serially searching the array until finding the target (Treisman & Gelade, 1980; Wolfe, 1994). Therefore, we expected response times in the Conjunction Search task to increase as a function of the number of distractor items present. Such endogenous control of attention is also required by our Flicker task, as participants must serially switch attention between items as they load items into VWM, make a comparison decision, and repeat the process until they find the target. For this reason we used search slope in the Conjunction Search task as a control for aspects of performance in the Flicker task that were related to controlled search, rather than to VWM capacity.

Children's performance in the Feature Search and Conjunction Search tasks is interesting in its own right (e.g., for estimating developmental changes in global visual processing, goal-directed search, and nonperceptual decision time) but is also of interest for our present purpose of controlling for these abilities when estimating developmental changes in VWM storage capacity.

Results and discussion

Descriptive statistics for each task, separated by age group, are presented in Tables 1 and 2. We present analyses of each task and relations between tasks below. All post-hoc comparisons were corrected using Scheffé's criterion. Because one of our primary interests was controlling for developmental changes in attention and executive control when evaluating possible developmental changes in VWM storage capacity, we present the results of the two visual search tasks first.

Feature search

First, as a check that participants were equally accurate across the different set sizes and to ask whether age was related to accuracy at locating a target defined by a single visual feature

Table 1 Descriptive statistics of measures across all tasks. Values in parentheses show the standard deviation from the mean.

Age (years)	Feature Search		Conjunction Search		Flicker	
	Average RT	Slope	Average RT	Slope	K	RT (non-search)
3	1.93 (0.47)	0.02 (0.04)	3.31 (0.72)	0.15 (0.10)	1.45 (0.53)	2.16 (0.73)
4	2.25 (0.37)	-0.02 (0.06)	3.19 (0.53)	0.12 (0.06)	1.83 (0.56)	2.33 (0.29)
5	1.96 (0.51)	0.01 (0.03)	2.67 (0.76)	0.09 (0.03)	1.95 (0.72)	2.13 (0.67)
6	1.96 (0.39)	0.01 (0.01)	2.52 (0.38)	0.07 (0.04)	2.96 (0.99)	2.06 (0.56)
7	1.68 (0.25)	0.01 (0.01)	2.53 (0.98)	0.06 (0.02)	2.94 (1.16)	2.04 (0.46)
8	1.54 (0.24)	0.00 (0.01)	2.25 (0.72)	0.05 (0.04)	3.90 (1.71)	1.79 (0.35)
Adults	0.75 (0.04)	0.00 (0.00)	1.08 (0.03)	0.06 (0.00)	4.00 (0.43)	1.05 (0.52)

(color), we ran a mixed model ANOVA of target identification accuracy (percent correct) with Age (3-, 4-, 5-, 6-, 7-, 8-year-olds, and adults) as a between-subjects factor and Distractor Set Size (2, 4, 8, and 12 items) as a within-subjects factor. This analysis yielded no effect of Age, $F(6,77) = 1.78$, $p = 0.11$, $\eta_G^2 = 0.06$, or Distractor Set Size, $F(3,231) = 0.57$, $p = 0.57$, $\eta_G^2 = 0.004$, nor any interaction of Age and Distractor Set Size, $F(18, 231) = 0.94$ $p = 0.51$, $\eta_G^2 = 0.04$. This suggests that, as expected, participants were similarly successful at locating the target of the feature search, regardless of age or number of distractors.

We next examined the more meaningful dependent measure of participants' response times. In a mixed model ANOVA of response times with Age (3-, 4-, 5-, 6-, 7-, 8-year-olds, and adults) as a between-subjects factor and Distractor Set Size (2, 4, 8, and 12 items) as a within-subjects factor, we found a main effect of Age, $F(6,77) = 22.02$, $p < 0.001$, $\eta_G^2 = 0.58$. Post-hoc contrasts revealed a significant difference in response times between adults and all other age groups combined, $F(1,77) = 91.03$, $p < 0.05$, as well as between adults and the oldest age group in our sample of children (8-year-olds), $F(1,77) = 24.17$, $p < 0.05$. There was no main effect of Distractor Set Size, $F(3,231) = 0.53$, $p = 0.64$, $\eta_G^2 = 0.001$.

The above main effect of Age does not solely reflect a difference between adults and children, as evidenced by a separate mixed model ANOVA conducted on children's response times only, with Age (3-, 4-, 5-, 6-, 7-, and 8-year-olds) as a between-subjects factor and Distractor Set Size (2, 4, 8, and 12 items) as a within-subjects factor. This analysis also revealed a main effect of Age, $F(5,66) = 5.00$, $p = 0.001$, $\eta_G^2 = 0.23$ (Fig. 2b). A post-hoc trend analysis showed a linear decrease in response times with Age, $F(1,66) = 15.97$, $p < 0.05$; hence, children's speed at finding the target improved over development. As predicted, we did not observe any effect of Distractor Set Size, $F(3,198) = 0.41$, $p = 0.72$, $\eta_G^2 = 0.001$, because response times were constant regardless of the number of distractors present. This pop-out effect was further reflected in analyses of response time slopes, which did not significantly differ from zero for any age group.[1] Furthermore, response time slopes were uncorrelated with Age in days, $r(70) = 0.02$, $p = 0.86$. However, the intercepts of the response time x set size regression functions were significantly correlated with Age in days, $r(70) = -0.37$, $p = 0.001$, replicating the result of the ANOVA reported earlier. These relations with Age can be seen in Table 3, which shows the simple correlations between the measures.

Overall, these data suggest developmental improvement in the Feature Search task (i.e., faster response times with age) between the ages of 3 and 8 years, as well as between 8 years and adulthood. These changes can be attributed to improvement in the speed of response exogenously to driven attention toward a singleton feature, improvements in global visual processing, decision-making, and executing a motor response.

Table 2 Accuracy rates (percent correct) for all tasks. Values in parentheses show the standard deviation from the mean

Age (years)	Feature Search	Conjunction Search	Flicker
3	98.75 (0.02)	99.00 (0.02)	94.92 (0.03)
4	100.00 (0.00)	100.00 (0.00)	96.83 (0.02)
5	99.25 (0.03)	99.75 (0.01)	99.00 (0.01)
6	99.75 (0.01)	100.00 (0.00)	99.67 (0.00)
7	100.00 (0.00)	100.00 (0.00)	100.00 (0.00)
8	100.00 (0.00)	99.75 (0.00)	99.33 (0.00)
Adults	100.00 (0.00)	100.00 (0.00)	98.96 (0.03)

Conjunction search

Across all ages, participants were equally successful at locating the target regardless of the number of distractors present. A mixed model ANOVA on target identification accuracy (percent correct) with Age (3-, 4-, 5-, 6-, 7-, 8-year olds, and adults) as a between-subjects factor and Distractor Set Size (2, 4, 8, and 12 items) as a within-subjects factor yielded no

[1] After correcting for multiple comparisons (all p-values >.01), slopes were not different from zero for any age group: 3-y-olds: $t(11)=1.68$, $p=0.12$; 4-y-olds: $t(11)=-1.51$, $p=0.16$; 5-y-olds: $t(11)=1.44$, $p = 0.18$; 6-y-olds: $t(11)=2.00$, $p = 0.07$; 7-y-olds: $t(11)=2.24$, $p = 0.05$; 8-y-olds: $t(11) = 0.432$, $p = 0.67$.

Table 3 Correlation matrix illustrating relationships (Pearson *r*) between measures collected from all tasks. Values in parentheses show significance level

	Age	Feature Search			Conjunction Search			Flicker Change Detection			
		Average RT	Slope	Intercept	Average RT	Slope	Intercept	Average RT	RT (non search)	K (Efficient)	K (Amnesic)
Age		-.43 (<.001)	.02 (.86)	-.37 (0.001)	-.49 (<.001)	-.52 (<.001)	-.27 (.02)	-57 (<.001)	-24 (.04)	.61 (<.001)	.61 (<.001)
Average RT			.09 (.44)	.88 (<.001)	.60 (<.001)	.10 (.39)	.64 (<.001)	.43 (<.001)	.48 (<.001)	-.34 (.004)	-33 (.005)
Slope				-.56 (<.001)	-.14 (.23)	.02 (.84)	-.14 (.23)	.03 (.78)	.06 (.64)	.01 (.94)	.01 (.95)
Intercept					.57 (<.001)	.08 (52)	.60 (<.001)	.33 (.004)	.36 (.002)	-28 (.02)	-28 (.02)
Average RT						.24 (.04)	.62 (<.001)	.55 (<.001)	.41 (<.001)	-.51 (<.001)	-.51 (<.001)
Slope							-.39 (.001)	.28 (.02)	.06 (.64)	-.41 (<.001)	-.41 (<.001)
Intercept								.44 (<.001)	.46 (<.001)	-.26 (.03)	-.25 (.03)
Average RT									.59 (<.001)	-.77 (<.001)	-.78 (.001)
RT (non search)										-30 (.01)	-30 (.01)
K (Efficient)											1.00 (<.001)
K (Amnesic)											

significant effect of Age $F(6,77) = 2.11$, $p = 0.06$, $\eta_G^2 = 0.04$ or Distractor Set Size, $F(3,231) = 1.45$, $p = 0.24$, $\eta_G^2 = 0.01$, nor any interaction of these $F(18,231) = 0.42$ $p = 0.93$, $\eta_G^2 = 0.02$, showing that all age groups were successful at finding the targets.

Next, we examined the more meaningful measure of response times, with Age (3-, 4-, 5-, 6-, 7-, 8-year-olds, and adults) as a between-subjects factor and Distractor Set Size (2, 4, 8, and 12 items) as a within-subjects factor. This yielded a main effect of Age, $F(6,77) = 25.46$, $p < 0.001$, $\eta_G^2 = 0.60$ (Fig. 2b). Post-hoc contrasts revealed significant differences in response times between adults and all other age groups combined, $F(1,77) = 82.64$, $p < 0.05$, as well as between adults and 8-year-olds, $F(1,77) = 17.11$, $p < 0.05$. As expected, we also observed a main effect of Distractor Set Size, $F(3,231) = 110.09$, $p < 0.001$, $\eta_G^2 = 0.26$. Post-hoc contrasts revealed differences in participants' reaction times between set sizes 2 and 4, $F(1,231) = 18.71$, $p < 0.05$, set sizes 4 and 8, $F(1, 231) = 90.52$, $p < 0.05$, and set sizes 8 and 12, $F(1, 231) = 29.26$, $p < 0.05$. Furthermore, the ANOVA yielded a significant Age x Distractor Set Size interaction, $F(18,231) = 3.36$, $p < 0.001$, $\eta_G^2 = 0.06$. A post-hoc contrast revealed that this effect was largely driven by a significant difference in response times observed at the smallest versus largest set sizes for 3-, 4-, and 5-year-olds versus 6-, 7-, and 8-year-olds and adults, $F(1,231) = 33.78$, $p < 0.05$. This suggests that older children and adults were less strongly affected by the presence of greater numbers of distractors.

These effects do not solely reflect differences between adults and children, as evidenced by an analysis of children only. An ANOVA on children's response times with Age (3-, 4-, 5-, 6-, 7-, and 8-year-olds) as a between-subjects factor and Distractor Set Size (2, 4, 8, and 12 items) as a within-subjects factor yielded a main effect of Age, $F(5,66) = 9.20$, $p < 0.001$, $\eta_G^2 = 0.35$ (Fig. 2b). A post-hoc trend analysis revealed a linear decrease in response times with age, $F(1,66) = 44.25$, $p < 0.05$, suggesting that children's ability to find the target—irrespective of set size—improved between ages 3 and 8 years. As predicted, we also observed a significant main effect of Distractor Set Size, $F(3,198) = 88.37$, $p < 0.001$, $\eta_G^2 = 0.24$. Post-hoc contrasts revealed significant differences in response times between set sizes 2 and 4, $F(1,198) = 27.28$, $p < 0.05$, set sizes 4 and 8, $F(1, 198) = 47.31$, $p < 0.05$, and set sizes 8 and 12, $F(1,198) = 24.19$, $p < 0.05$. Hence, as expected, having more distractors present slowed children's responses. We also observed a significant interaction of Age and Distractor Set Size, $F(15,198) = 3.26$, $p < 0.001$, $\eta_G^2 = 0.06$. Post-hoc analyses suggest that this effect was primarily due to younger children showing a greater difference in reaction time between the smallest and the largest set sizes than the other age groups, $F(1,198) = 35.83$, $p < 0.05$.

Next, to better understand the relation between age and Conjunction Search speed, we performed a one-way ANOVA on the slope of response times across distractor set sizes, with Age (3-, 4-, 5-, 6-, 7-, 8-year-olds, and adults) as the between-subjects variable. This revealed a significant effect of Age, $F(6,77) = 5.80$, $p < 0.001$, $\eta_G^2 = 0.31$. However, post-hoc analyses revealed that this effect was not driven by differences in search rates between adults and all children, $F(1,77) = 2.22$, $p > 0.05$, nor between adults and 8-year-olds, $F(1,77) = 0.45$, $p > 0.05$. When we re-ran the ANOVA excluding adults, we found a significant effect of age, $F(5,66) = 5.53$, $p < 0.001$, $\eta_G^2 = 0.30$. A post-hoc trend analysis revealed a significant decrease in search slope with age, $F(1,66) = 27.03$, $p < 0.05$. This developmental improvement was further reflected in a significant correlation between Conjunction Search slopes and Age in days, $r(70) = -0.52$, $p < 0.001$. Hence, children's speed at serially searching a visual array to locate a feature conjunction improved with age.

The intercept of the Reaction Time x Set Size functions also significantly correlated with Age in days, $r(70) = -0.27$, $p = 0.02$, suggesting improvements in the non-search components of the task. Although these intercept values significantly correlated with those in the Feature Search task, $r(70) = 0.60$, $p < 0.001$, Conjunction Search intercepts were slightly higher, $t(71) = -1.92$, $p = 0.06$. This marginal difference may have resulted from the more visually complex displays shown in the Conjunction Search task.

Overall, these data suggest developmental improvement in the Conjunction Search task (i.e., faster response times and decreased search slopes) between the ages of 3 and 8 years, as well as between 8 years and adulthood. These changes are likely due to improvements in the executive control processes involved in the endogenous control of attention and in the ability to search an array for a complex visual target.

Flicker change detection

We began our analyses of performance in the Flicker task by checking that participants' success at eventually locating the changing target did not vary by age. A one-way between-subjects ANOVA on target identification accuracy (percent correct averaged for set sizes 4 and 6) with Age (3-, 4-, 5-, 6-, 7-, 8-year-olds, and adults), revealed no significant effect of Age, $F(6,77) = 1.80$, $p = 0.11$, $\eta_G^2 = 0.12$. All age groups were equally successful in finding the target.

Next, we examined the more meaningful measure of response times. To create a single measure from the separate RTs from set sizes 4 and 6, we investigated the possibility of using the average RT across all trials. We found this measure to be normally distributed for each age group (Fig. 4), and therefore we conducted subsequent analyses using this combined measure.

First, we conducted a one-way ANOVA on average RT with Age (3-, 4-, 5-, 6-, 7-, 8-year-olds, and adults) as a between-subjects factor. This yielded a significant effect of Age, $F(6,77) = 13.29$, $p < 0.001$, $\eta_G^2 = 0.51$. To further explore this developmental change, we performed a linear regression on Age (in days) and average RT. This revealed a significant effect $F(1,70) = 34.07$, $p < 0.001$, with a negative slope suggesting a decrease of approximately 365 msec of average search time per year. This replicates the effects of the ANOVA and again reveals improving search rate in the Flicker task with age.

Developmental changes in children's VWM, controlling for perceptual, attentional, and executive control factors

Given that RT in our Flicker task correlated with age, as did performance in our two visual search tasks (Table 3), it is important to address the possibility that the observed developmental improvement in Flicker performance might have resulted from changes in attentional control, global visual processing, or executive control abilities, rather than changes in VWM capacity itself.

To address this, we first examined the relation between age and RT in the Flicker task, regressing out the contribution of average response time in the Feature Search task (i.e., the RT involved in the efficient search performed when parallel processing was possible, and when attention was exogenously driven by the presence of a singleton feature). In addition, we regressed out search slope in the Conjunction Search task (i.e., the RT involved in the inefficient search performed when serial processing was required, and when attention was endogenously moved serially within the array). We performed two separate linear regressions to create residuals for Flicker average RT and Age while controlling for Feature Search average RT and Conjunction Search slope. Correlating the residuals revealed a significant relationship between RT in the Flicker task and Age (in days) across childhood, $r(70) = -0.41$, $p < 0.001$. This shows that there are developmental improvements in average RT in the Flicker task that remain robust when developmental improvements in attention, global processing, and executive control are accounted for.

An even more robust control would be to estimate and control for nonsearch-related processes that may have occurred within the Flicker task itself. For example, participants might wait an extra cycle to confirm that they saw a change at the target location—a response delay that would add to our estimate of visual search, or participants might watch the first few cycles of the flickering screen as they gather information about the general layout of the array and plan their search—a strategy that also would add to response time. To the extent that participants engaged in these and other activities not directly related to visual search, our measure of active search will be distorted. Furthermore, if the tendency to engage in

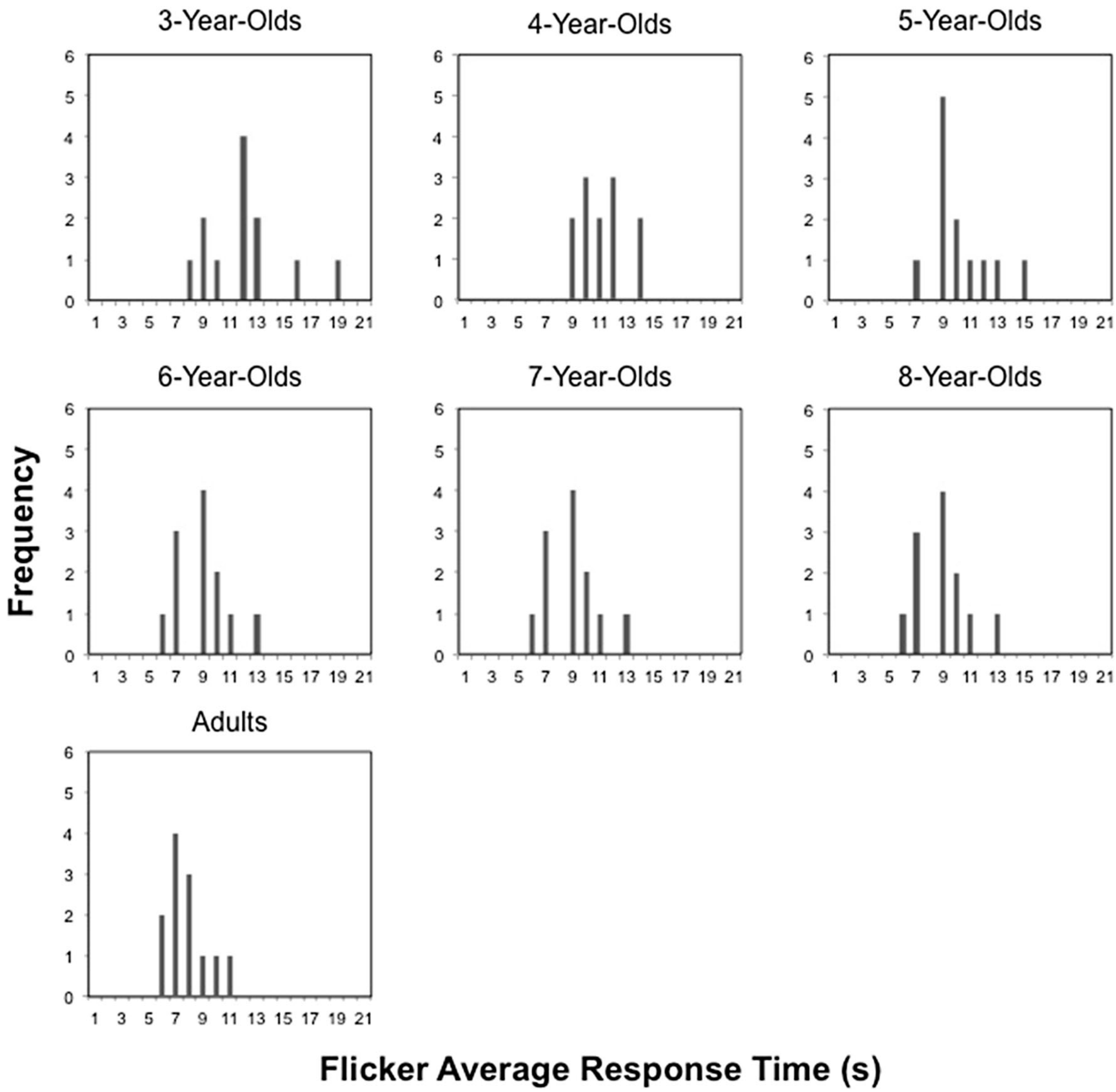


Fig. 4 Histograms of Flicker response time (s) averaged across set sizes for each age group

these non-search activities changes with age, observed developmental differences in performance will be affected. To take these factors into account, we attempted to estimate these processes from performance in the Flicker task itself.

To do this, we first rank ordered participants' response times (Fig. 5), and then, using these rank orders, identified the average of the three fastest trials as an estimate of nonsearch-related processing (Pailian & Halberda, 2015). These trials, in which the target happened to be located quickly and therefore the duration of active search was minimal, reflect whatever time children took to look at the screen before initiating their search, as well as the time they waited to verify the changing target before making a response. Figure 5 shows that the average of the three fastest RTs appears to be roughly a midpoint between the first fastest trial and the first five fastest trials—we find that any choice in this range produces similar results.

A one-way ANOVA on the Flicker non-search related activity (i.e., the average of the three fastest correct RTs) with Age in years (3-, 4-, 5-, 6-, 7-, 8-year-olds, and adults) as a between-subjects factor produced a significant effect, $F(6,77) = 7.61$ $p < 0.001$, $\eta_G^2 = 0.37$ (Fig. 6). Post-hoc contrasts revealed that Flicker nonsearch-related activity for adults did not significantly differ from that for 8-year-olds, $F(1,77) = 11.47$, $p > 0.05$. However, a significant difference in Flicker non-search related activity was observed between adults and 7-year-old children, $F(1,77) = 20.60$, $p < 0.05$. This suggests that Flicker nonsearch-related activity may be somewhat different between older children and adults.

Next, we included Flicker non-search related activity in our test for developmental improvement in Flicker performance in childhood. We used two separate linear regressions to create residuals for Flicker average RT and Age in days that were controlled for Flicker nonsearch-related activity,

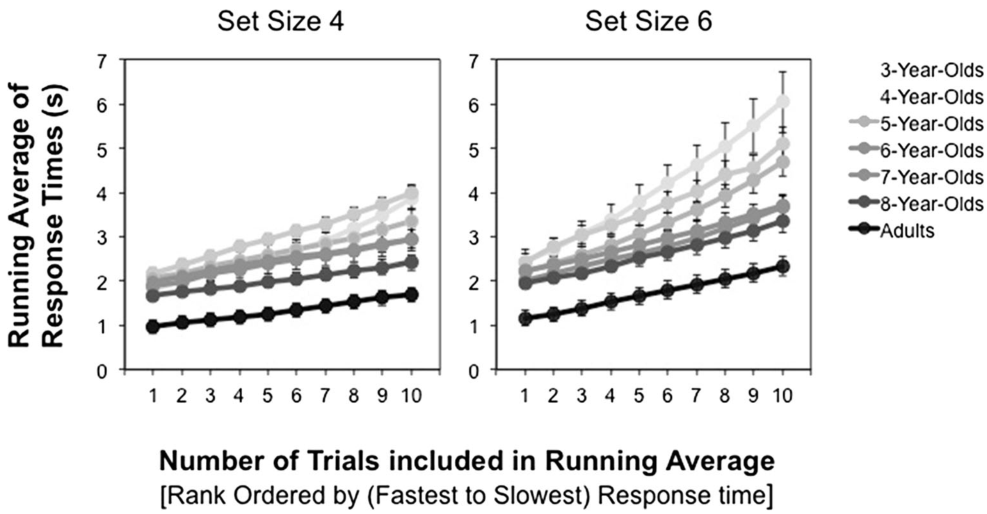


Fig. 5 Running average of Flicker response time (s) as a function of rank order based on the x many fastest trials

Feature Search average RT, and Conjunction Search slope. The correlation between the resulting residuals was significant, $r(70) = -0.45$, $p < 0.001$.[2] These analyses suggest that there are improvements in VWM performance during childhood, even when controlling for developmental improvements in visual processing and decision making during tasks that tap exogenously controlled attention (i.e., Feature Search), and tasks that tap active serial visual search for a target (i.e., Conjunction Search), and when accounting for nonsearch-related activity within the Flicker task itself (i.e., average of three fastest RTs).

Estimating VWM storage capacity (K) from response time in the Flicker task

Our principal measure in the Flicker task was response time, but we also explored a method for transforming RTs into separate estimates of VWM storage capacity and nonstorage-related factors that may additionally influence RTs, such as search efficiency, display duration, and nonsearch-related activity (Pailian & Halberda, 2015). We provide the results of this exploratory analysis, with the aim of determining a range of possible VWM capacities from our data.

Estimating the storage capacity of VWM using RTs is possible because a participant's rate to search for the target item is directly affected by their VWM storage capacity. For example, with a VWM storage capacity of two items, a participant might maintain information from two of the items in the array in VWM over the blank interval, then compare these remembered items to the items presented in the next array. With a VWM storage capacity of only one item, a different participant might maintain just a single item during the blank interval. In this way, participants with larger VWM storage capacity will be able to search the array more quickly, yielding shorter average RTs. Hence average RT can be used to estimate average VWM storage capacity for each participant.[3]

Second, RT on a given trial in the Flicker task will be affected by the number of items a participant searches before finding the changing target. Specifically, RTs will be faster when visual search is maximally "efficient" (random without replacement). An example of efficient search would be when a participant looks at each stimulus item for the minimum amount of time required to compare the contents of VWM with the observed array, and also never reexamines a previously visited item. Previous modeling work on visual search has revealed that, if search is efficient, then a participant will search, on average, (set size + 1) /2 items before finding the target (Johnson & Kotz, 1977; Treisman & Gelade, 1980). Conversely, RTs will be longer when search is "inefficient" (Horowitz & Wolfe, 1998, 2003)—for example when a participant does not store information about which items have already been searched. If search is inefficient, then on average a participant must search a number of items that is equal to the total number of items in the array before finding the target (Johnson & Kotz, 1977; Kontsevich, 2001). Given that participants may adopt a variety of strategies representing a mix of efficient and inefficient search (e.g.,

[2] To be conservative, we also performed a similar linear regression in which we controlled for all possible measures (Feature Search slope, Feature intercept, Feature average RT, Conjunction Search slope, Conjunction intercept, Conjunction average RT, and Flicker non-search related activity), and found that the relation between Flicker average RT and Age in days remained significant ($r = 0.-.36$, $p = 0.004$).

[3] We note that this approach remains essentially unchanged if one considers VWM to be limited by total information rather than by the number of items (see Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Brady, Konkle, & Alvarez, 2009; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). Our interest in using RT in the Flicker task to estimate VWM storage capacity is neutral with respect to this important debate. Throughout, we will refer to an object-limited VWM, but our claims also translate into units of information rather than objects.

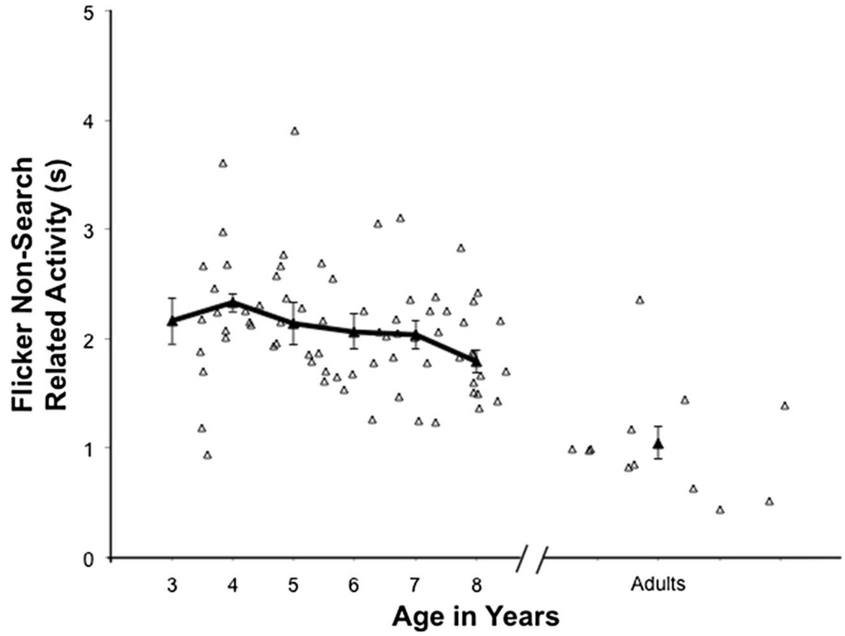


Fig. 6 Nonsearch-related activity (s) in the Flicker task (estimates of executive control abilities), based on average of three fastest RTs across all set sizes. Filled symbols plot group means and error bars show the standard error for each mean value. Unfilled symbols plot nonsearch-related activity for individual participants

spending an additional cycle on certain items, or repeatedly visiting particularly visually salient item), we provide a range of estimates that represent the lower and upper bounds of VWM storage capacity that result from search being either maximally efficient (random without replacement) or inefficient (random with replacement).

Third, nonsearch-related activity may distort search time in the Flicker task. Because VWM storage capacity should be estimated by a participant's rate to search actively the display, and not their rate to plan their search or verify their answer after target detection, it is important to factor out non-search related activity when estimating VWM storage capacity. Earlier we adopted the approach of using the average of each participant's three fastest correct response times to estimate non-search related activity in the Flicker task. By subtracting this estimate of non-search related activity from the average RT, we can better estimate the amount of time a participant spends actively searching the display.[4]

As a final consideration, search rate in the Flicker task is partially determined by the rate of alternations of the displays themselves, because participants must compare items from one flash to the next (i.e., search rate = display duration + blank duration).

Below, we present formulas for two possible modes of search efficiency that translate average response time in the Flicker task (RT) into an estimate of VWM storage capacity. The symbol (K) designates estimated VWM storage capacity.

[4] This approach is still being refined, however, elsewhere we have explored the benefits of this same approach in adults and found it valuable (Pailian & Halberda, 2015).

Equation 1a (efficient search):

$$\mathrm{RT}_{average} - \mathrm{RT}_{non-search} = (\text{display duration} + \text{blank duration}) * \frac{[(\text{set size} + 1)/2]}{\mathrm{K}}$$

Equation 1b (inefficient search):

$$\mathrm{RT}_{average} - \mathrm{RT}_{non-search} = (\text{display duration} + \text{blank duration}) * \frac{(\text{set size})}{\mathrm{K}}$$

On the left side of both equations is an estimate of the average amount of time a participant spends actively searching the display ($\mathrm{RT}_{average} - \mathrm{RT}_{non\text{-}search}$). On the right side is a specification of how search RT emerges from search rate (display duration + blank duration) and the average number of items searched before finding the target. In each equation, the number of items searched is divided by the participant's VWM storage capacity (K). Intuitively, the VWM storage capacity (K) is serving to "chunk" the total number of items that have to be searched.

Rearranging these equations yields a formula for estimating VWM storage capacity (K):

Equation 1c (efficient search):

$$\mathrm{K} = \frac{(\text{display duration} + \text{blank duration}) * [(\text{set size} + 1)/2]}{\mathrm{RT}_{average} - \mathrm{RT}_{non-search}}$$

Equation 1d (inefficient search):

$$K = \frac{(\text{display duration} + \text{blank duration})*(\text{set size})}{RT_{average} - RT_{non-search}}$$

Using Equations 1c and 1d, we estimated K for each set size, for each participant. To obtain an estimate of VWM storage capacity that was unaffected by outliers, response times (RT) greater than two standard deviations from the mean were removed for each participant before calculating K (4% of trials).

Because the time spent on nonsearch-related activity might change with set size, we calculated a separate estimate of $RT_{non\text{-}search}$ for each set size (e.g., $RT_{non\text{-}search}$ for set size 4 was calculated using the average of the three fastest RTs at set size 4) .[5] We also calculated average response time ($RT_{average}$) for each set size. From these two estimates (i.e., $RT_{average}$ and $RT_{non\text{-}search}$) and the remaining fixed values (e.g., display duration, blank duration, set size) we computed estimates of VWM storage capacity (K) for each participant, for each set size, for both types of search efficiency (Equations 1c and 1d).

The resulting K estimates proved reliable: for both types of efficiency (i.e., efficient and inefficient search), K values calculated using set sizes 4 were significantly correlated with K values calculated using set size 6 for both children and adults (efficient search, $r(82) = 0.65$, $p < 0.001$; inefficient search, $r(82) = 0.65$, $p < 0.001$). Such correlations might emerge simply as a result of older children and adults having larger K values than younger children; we therefore examined children's performance while controlling for developmental changes in VWM storage capacity. We first performed linear regressions of K with age in days for each set size and each type of search efficiency. This created residual estimates of individual differences, controlled for age. We then performed a linear regression between K set size 4 and K set size 6 for each type of search efficiency and found that individual differences at K set size 4 related to individual differences at K set size 6 controlling for age (efficient search, $r(70) = 0.55$, $p < 0.001$; inefficient search $r(70) = 0.55$, $p < 0.001$).

Because of the relation between K at set sizes 4 and 6, we next averaged K values across set sizes 4 and 6 to produce an overall estimate of each participant's VWM storage capacity for both types of search efficiency. The resulting estimates can be seen in Fig. 7, which shows that VWM storage capacity (K), as derived from response time on the Flicker task, improves throughout childhood. The range of capacities (K) are consistent with developmental improvements in VWM storage capacity (K), even if search strategies change with age. That is, even if 3-year-olds adopt an inefficient search strategy (resulting in their true VWM storage capacity falling in the higher part of the range for their age) and 8-year-olds adopt a more efficient search strategy (resulting in their true VWM storage capacity falling in the lower part of the range for their age), there is still significant developmental improvement in VWM storage capacity (K) (note also the positive trend across time in Fig. 7).

Next, we performed a one-way between-subjects ANOVA on averaged Flicker K estimates assuming inefficient search, with Age in years (3-, 4-, 5-, 6-, 7-, 8-year-olds, and adults) as the between-subjects variable.[6] This revealed a significant effect of Age, $F(6,77) = 9.94$, $p < 0.001$, $\eta_G^2 = 0.44$. Post-hoc contrasts revealed no significant differences between the ages of 3 to 5 years (Ages 3 vs. 4: $F(1,77) = 0.74$, $p > 0.05$ Ages 3 vs. 5: $F(1,77) = 1.30$, $p > 0.05$; Ages 4 vs. 5: $F(1,77) = 0.08$, $p > 0.05$) nor between ages 6 to 8 years (Ages 6 vs. 7: $F(1,77) = 0.001$, $p > 0.05$; Ages 6 vs. 8: $F(1,77) = 4.50$, $p > 0.05$; Ages 7 vs. 8: $F(1,77) = 4.66$, $p > 0.05$). Furthermore, post-hoc contrasts revealed no significant differences in Flicker K estimates between adults and 8-year-olds, $F(1,77) = 0.05$, $p > 0.05$, adults and 7-year-olds, $F(1,77) = 5.69$, $p > 0.05$, or adults and 6-year-olds, $F(1,77) = 5.52$, $p > 0.05$. However, a significant difference was observed when comparing Flicker K estimates between adults and preschoolers (adults vs. 5-year-olds: $F(1,77) = 21.50$, $p < 0.05$; adults vs. 4-year-olds: $F(1,77) = 24.19$, $p < 0.05$; adults vs. 3-year-olds: $F(1,77) = 33.38$, $p < 0.05$). Lastly, a post-hoc contrast comparing Flicker K estimates for 3- to 5-year-old children versus 6- to 8-year-old children and adults yielded a significant difference, $F(1,77) = 51.37$, $p < 0.05$.

Taken together, these exploratory analyses suggest that VWM capacity (K) improves during early childhood and bears resemblance to that of adults by around 8 years of age. These results on VWM capacity (K) provide a complimentary picture to the one that emerged from the analysis of Flicker average RT.

General discussion

Our results suggest that Visual Working Memory (VWM) storage capacity improves between ages 3 and 8 years and approaches adult-like levels between 6 and 8 years (although examining possible developmental change between 8 years and adulthood would be additionally valuable). Using a Flicker change detection task designed for children, along with Feature and Conjunction Search tasks, we measured

[5] Note also that this makes it inadvisable to estimate K from the slope between set size and RT as it assumes a common intercept across set sizes, leaving all increase in RT to be interpreted as increasing search time. In fact some of this increase may result from longer non-search related activity prior to and after search with some set sizes more than others. , It is therefore prudent to estimate non-search related activity separately for each set size.

[6] Note that the pattern of results remains statistically identical whether search is assumed to be inefficient or efficient, given that estimates based on these models differ by a constant ratio.

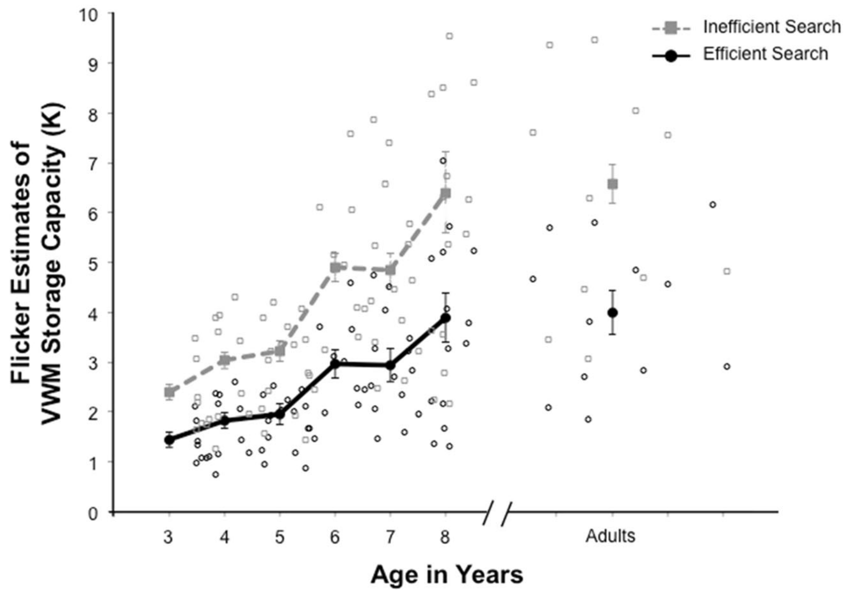


Fig. 7 Estimates of VWM storage capacity (K), observed in our Flicker task based on assumed search efficiency. Filled symbols plot group means and error bars show the standard error for each mean value. Unfilled symbols plot capacity estimates for individual participants

VWM capacity across a range of ages and controlled for developmental improvements in attention, global visual processing, and executive control. We found that the amount of time it took children to find the target in both the Feature and Conjunction Search task decreased with age, as did the amount of time needed for decision-making and motor response execution within the Feature Search task. The impact of distractors on search rate during Conjunction Search also decreased with age. These findings suggest that there are developmental improvements in the abilities that support performance in these search tasks. Critically, none of these accounted for the observed improvements in VWM performance on the Flicker task. When we removed the contribution of Feature and Conjunction Search performance from children's performance on the Flicker task, we found that the performance correlation with age remained. Furthermore, this was true when we controlled for nonsearch-related processes within the Flicker task itself. This suggests robust improvements in VWM performance during the childhood years.

One of our motivations for adapting the Flicker task for use with children was our recent work with adults, which measured performance in the Flicker task alongside the more traditionally used One-Shot paradigm. Importantly, the reliability and stability of these paradigms for measuring VWM capacity had not previously been assessed. We found that in adults, the One-Shot paradigm, as typically used in many leading papers, was less reliable than the Flicker paradigm in returning a stable measure of VWM capacity (Pailian & Halberda, 2015). Indeed, performance in the One-Shot paradigm, as typically used, did not even correlate with *itself* across set sizes (Pailian & Halberda, 2015). Based on our work, it appears that the One-Shot task measures different aspects of processing at different set sizes. In ongoing work we are attempting to characterize the other abilities tapped by the One-Shot task (e.g., ensemble coding is one possibility). Our interest was not so much to compare the One-Shot task to the Flicker task as an estimate of VWM capacity (which is a focus of our other work: Pailian & Halberda, 2015; Halberda, Pailian, Wetherhold, & Simons, unpublished data). Rather, our approach was to try to measure the development of visual abilities across several tasks (drawing on perception, attention, and executive control) and to control for these when analyzing performance in a VWM task (e.g., Flicker). Measuring and controlling for a variety of abilities, rather than taking any one task (e.g., One-Shot) to be a "gold standard" that unambiguously measures a particular psychological construct, may be a richer approach for understanding development and change

The low reliability of the One-Shot task may be one reason for the variable results of previous studies using this method in children (Cowan et al., 2005; Simmering, 2012; Riggs et al., 2006). Additionally, in adults, the One-Shot paradigm has been shown to engage diffuse attention to many items (e.g., ensemble features) rather than engaging purely object-based attention (Alvarez, 2011; Brady, & Alvarez, 2015; Brady & Tenenbaum, 2013). However, rather than critiquing any particular version of the One-Shot task, or undermining its ability to measure VWM capacity, we emphasize that tasks with very different parameters and goals can provide converging evidence for VWM capacity across development. It will be important to test constructs, such as VWM, using a range of paradigms and critically to consider (and control for) the various types of processing invoked by each.

Our results are consistent with a period of protracted development of VWM capacity during early childhood, along with improvements in perception, attention, and executive control. Because the Flicker task is a relatively new measure of VWM capacity, continued evaluation of this task seems required. Questions remain, including how best to translate RT into an estimate of VWM capacity, how search should be characterized within the task (e.g., is it efficient or inefficient), how one can estimate nonstorage-related components of the task, and what are the contributions of storage, comparison, the purging of information from VWM, and uploading new information into VWM. We hope that this new paradigm can be employed with both children and adults to address these open questions.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131.

Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science, 15*(2), 106–111.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences USA, 108,* 10367–10371.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items, regardless of complexity. *Psychological Science, 18*(7), 622–628.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Barbot, A., Landy, M. S., & Carrasco, M. (2012). Differential effects of exogenous and endogenous attention on 2nd-order texture contrast sensitivity. *Journal of Vision, 12*(8), 6, 1–15.

Becker, M. W., Pashler, H., & Anstis, S. M. (2000). The role of iconic memory in change- detection tasks. *Perception, 29,* 273–286.

Brady, T. F., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*(3), 921–929.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*(4), 487–502.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and towards structured representations. *Journal of Vision, 11*(5), 4, 1–34.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher-order regularities into working memory capacity estimates. *Psychological Review, 120*(1), 85–109.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology, 17,* 177–184.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioural and Brain Sciences, 24,* 87–185.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51,* 42–100.

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition, 34*(8), 1754–1768.

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science, 13,* 120–133.

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: Not based on encoding limitations. *Developmental Science, 14,* 1066–1074.

Donderi, D. C., & Zelnicker, D. (1969). Parallel processing in visual same-different decisions. *Perception & Psychophysics, 5,* 197–200.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–358.

Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology, 3,* 674–698.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97,* 295–313.

Flombaum, J. I., & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception & Performance, 32*(4), 840–853.

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological Science, 22*(3), 361–368.

Fukuda, K., Vogel, E. K., Mayr, U., & Awh, E. (2010). Quantity not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin and Review, 17,* 673–679.

Gerhardstein, P., & Rovee-Collier, C. (2002). The development of visual search in infants and very young children. *Journal of Experimental Child Psychology, 81*(2), 194–215.

Halberda, J., Pailian, H., Wetherhold, J., & Simons, D. (2006). *You can never attend to more than three items at once: Gestalt grouping principles explain changes in capacity*. Manuscript in preparation.

Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distractors in visual search? *Visual Cognition, 10*(3), 257–298.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual Search Has No Memory. *Nature, 357,* 575–577.

Johnson, S. J., Hollingworth, A., & Luck, S. J. (2008). The role of attention in the maintenance of feature bindings in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 41–55.

Johnson, N. L., & Kotz, S. (1977). *Urn models and their applications*. New York: John Wiley & Sons.

Johnston, J. C., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 843–856.

Kahneman, D., Treisman, A. M., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Kontsevich, L.L (2001). *Visual search is certainly serial. Paper presented at the meeting of the Vision Sciences Society, Saraosta, FL.*

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.

Oakes, L. M., Messenger, I. M., Ross-Sheehy, S., & Luck, S. J. (2009). New evidence for rapid development of color-location binding in infants' visual short-term memory. *Visual Cognition, 17,* 67–72.

Pailian, H., & Halberda. (2015). The reliability and internal consistency of one-shot and flicker change detection for measuring individual differences in visual working memory capacity. *Memory & Cognition, 43*(3), 397–420.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44,* 369–378.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Percept. Psychophys., 16,* 283–290.

Quinlan, P. T. (2003). Visual feature integration theory: past, present, and future. *Psychological bulletin, 129*(5), 643.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373.

Riggs, K. J., McTaggart, J., Simpson, A., & Freeman, R. P. J. (2006). Changes in the capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology, 95,* 18–26.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74,* 1807–1822.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences (PNAS), 105*(16), 5975–5979.

Scholl, B. J., & Xu, Y. (2001). The magical number 4 in vision. [Commentary]. *Behavioral and Brain Sciences, 24*(1), 145–146.

Shinskey, J. L., & Munakata, Y. (2003). Are infants in the dark about hidden objects? *Developmental Science, 6,* 273–282.

Simmering, V. R. (2012). The development of visual working memory capacity during early childhood. *Journal of Experimental Child Psychology, 111,* 695–707.

Simmering, V. R., & Perone, S. (2013). Working memory capacity as a dynamic process. *Frontiers in Developmental Psychology, 3,* 567.

Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr., 74,* 1–29.

Theeuwes, J. (1993). *Endogenous and exogenous control of visual selection: A review of the literature*. Retrieved from DTIC database. http://www.dtic.mil/dtic/tr/fulltext/u2/a273761.pdf.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Vogel, E., & Awh, E. (2008). How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. *Current Directions in Psychological Science, 17,* 171–176.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures of individual differences in controlling access to working memory. *Nature, 438*(24), 500–503.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance, 27,* 92–114.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1436–1451.

Wolfe, J. M. (1994). Guided Search 2.0: A Revised Model of Visual Search. *Psychonomic Bulletin & Review, 1*(2), 202–238.

Wolfe, J. M. (1998). What Can 1,000,000 Trials Tell Us About Visual Search? *Psychological Science, 9,* 33–39.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433.

Wolfe, J. M., Oliva, A., Butcher, S. J., & Arsenio, H. C. (2002). An unbinding problem? The disintegration of visible, previously attended objects does not attract attention. *Journal of Vision, 2*(3), 256–271.