

Development of the ability to combine visual and acoustic information in working memory

Nelson Cowan | Yu Li | Bret A. Glass | J. Scott Saults

Department of Psychological Science, University of Missouri, Columbia, MO, USA

Correspondence

Nelson Cowan, Department of Psychological Science, University of Missouri, McAlester Hall, Columbia, MO 65211, USA.
Email CowanN@Missouri.edu

Funding information

NIH Grant R01 HD-21338

Abstract

Presentation of two kinds of materials in working memory (visual and acoustic), with the requirement to attend to one or both modalities, poses an interesting case for working memory development because competing predictions can be formulated. In two experiments, we assessed such predictions with children 7–13 years old and adults. With development, the ability to hold more information in the focus of attention could lead to an increase in the size of the trade-off between modalities; if attention can hold A items during unimodal-attention trials, then on average attention should hold $A/2$ of those same items during bimodal-attention trials. If A increases with age, so would the dual-task cost, $A/2$. The results clearly ruled out that possibility. It was the modality- or code-specific components of working memory that improved with age and not the central component. We discuss various mechanisms that could have produced these results, including alternative attention-based mechanisms. The findings point to a rich field for continued research.

RESEARCH HIGHLIGHTS

- Attention processes in working memory development were examined with acoustic lists and visual arrays, the requirement being retention of sometimes one and sometimes both sets.
- Comparison of trials requiring unimodal versus bimodal attention yielded estimates of three faculties: one dedicated to acoustics, one to visuals, and one shared between them.
- Competing hypotheses produced contrasting expectations: capacity could develop because of an increasing ability to share attention between modalities or increasing ability to hold visual and acoustic information without shared attention.
- With colors and either spoken digits (Experiment 1) or tones (Experiment 2), separate visual and acoustic memory storage developed from 7 years to adulthood; interestingly, attention-based sharing stagnated.

1 | INTRODUCTION

Working memory refers to the small amount of information held in a temporarily heightened state of availability (Baddeley & Hitch, 1974; Cowan, 2017), and performance on tasks measuring working memory increases markedly throughout childhood (Gathercole, Pickering, Ambridge, & Wearing, 2004). After many years of investigation, however, the fundamental reasons for increases in working memory across development remain unclear. Working memory is likely to be critical for understanding many aspects of development. In conceptual development, for example, understanding the basic concept of a tiger requires realization that it is a big, striped cat. A different concept results if any one feature is forgotten (e.g., not-small, house cat; not striped, lion; or not a cat, zebra). With maturity, higher-capacity working memory could allow more complexity of concepts (Halford, Cowan, & Andrews, 2007). Working memory is needed to retain information as it is being processed, in problem-solving and language use. Given that working memory is involved in many developmental disorders (e.g., see Cowan & Alloway, 2009; Henry, 2012), it is important to

understand the mechanisms of working memory for practical as well as theoretical reasons.

For several decades, the predominant view among developmental researchers was that the capacity of working memory, or number of items retained concurrently, did not necessarily change with development. There were cogent arguments for the roles of other processes, including the developing speed of various processes (Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Case, Kurland, & Goldberg, 1982; Hulme & Tordoff, 1989; Kail & Park, 1994), knowledge (Chi, 1978), mnemonic strategies (Flavell, Beach, & Chinsky, 1966; Ornstein & Naus, 1978), and the ability to exclude distractions (Hagen, 1967). Cowan (2016) reviewed more recent evidence that these additional factors, although clearly important, do not seem to account fully for the development of working memory; capacity, or an attentional resource underlying it, could be a primary developing factor after all (cf. Case, 1995; Pascual-Leone & Johnson, 2011; Pascual-Leone & Smith, 1969). Still, it remains unclear whether capacity increases as a fundamental aspect of working memory development or if, instead, other basic processes underlie the increases that appear to reflect capacity development.

1.1 | The developmental question and background in adult research

The present work investigates whether working memory develops primarily because more information can be held using attention, or whether it develops primarily because more information can be held in a way that does not require as much attention for maintenance.

In adults, Cowan, Sauls, and Blume (2014) were able to investigate attentional demands of working memory by presenting two sets of stimuli, an array of colored spots and a series of verbal items, in either order within a trial. In a given trial block, participants were to attend to and remember only the visual items or only the verbal items (i.e., two kinds of unimodal trial blocks, the visual-load block and the auditory-load block), or both sets together (bimodal-load trial blocks). There was a recognition test for one item at the end of each trial. In order to maximize the possibility that both sets would have to draw upon a common mental resource if one exists, articulatory suppression, or repetition of a word during the working memory task, was used. Suppression would prevent the participants from using covert verbal rehearsal to retain the verbal set in a form that might not require much attention (Guttentag, 1984; Morey & Cowan, 2004). Under these circumstances, it was possible to dissect performance into three types of mnemonic faculty: one dedicated consistently to the visual items regardless of the attention condition (termed a peripheral visual component); one dedicated consistently to the verbal items (termed a peripheral verbal component); and one that, depending on the attention condition, could enhance visual memory, verbal memory, or some of each (termed a central component). Across a number of experiments, memory in the visual-alone unimodal attention condition totaled about 3.5 items on average (theoretically, 2.5 from peripheral visual memory plus 1 more retained in an attention-based resource) whereas, when both sets had to be attended, in the

bimodal condition, this visual performance was reduced by about half an item on average (presumably, 2.5 peripheral visual memory + 0.5 from the divided central resource). Verbal performance similarly depended on the attention conditions, with similar levels of performance. In developmental research using this technique, we can determine which components of working memory increase in capacity with development.

1.2 | Developmental adaptation

The present work adapts the procedure of Cowan et al. (2014) for developmental study of children from the early elementary school years through adulthood. It is possible to imagine several strikingly different ways in which the pattern just described could change with development. Given that both visual and acoustic working memory increase with age (e.g., Cowan, 2016; Gathercole et al., 2004), it could be that the peripheral components increase with age; it could be that the central component increases; or it could be that both central and peripheral components increase with age. All of these are reasonable possibilities. In favor of development of the peripheral components, for example, participants with more life experiences might be more often reminded of something they know when random patterns or sequences are presented, resulting in chunking (Miller, 1956) and, therefore better memory. In favor of the development of the central component, it is well known that attentional abilities do increase with age (Ristic & Enns, 2015), and this could result in more items held using attention.

The fate of the central component with development depends, though, on just how attention is used when it must be flexibly allocated to assist performance in trials with two stimulus sets, under both unimodal and bimodal instructions. When the first set of stimuli is presented on a trial, more mature participants could be better at recoding this set in a manner that relies less on attention. For example, Cowan et al. (2014) suggested that information is off-loaded from a component of working memory termed the focus of attention to another component termed the currently activated portion of long-term memory (according to a theory proposed by Cowan, 1988, 1999, 2005). In the bimodal trials, off-loading the first set would make the focus of attention more readily available for the encoding of a second set into working memory, which would reduce the degree to which attention had to be flexibly allocated to one modality or the other. Thus, with development, the conflict between visual and verbal materials in working memory would be reduced by the off-loading process. The developmental course of the central component theoretically could depend on whether the predominant factor is how much information can be held in attention (potentially increasing with maturity) or on how well stimulus processing can be used to avoid holding information in attention (potentially decreasing storage in attention with maturity or at least leaving it constant, despite increases in performance). Below, we present the background behind the measurement of working memory in a way that allowed assessment of central and peripheral components to determine which developmental trends take place.

1.3 | Quantification of working memory components

One reason why this work can be carried out is that there is a simple, yet useful way to estimate the number of items held by each theoretical component of working memory (Cowan et al., 2014). The first step of this process is to estimate the number of items held in working memory in a particular condition. Suppose that k items are held in working memory and the set to be remembered includes N items. In the memory test, a single item will be presented at the center of the screen and must be judged part of the set or not included in that set. The task must be carried out by trying to match this probe item to each item in the set. If the probe was in the set, the likelihood that it is present in the participant's memory, allowing a match to be found, is k/N . If the item was in the set but is not in working memory, or if the item was not in the set, the outcome is that the participant can only guess at the answer. With some likelihood g , the participant will guess that the item was in the memory set. It can be shown that $k = N^*(h-f)/h$, where h is the probability of a hit, defined as a correct indication that the probe item is new and not in the set to be remembered, and f is the probability of a false alarm or incorrect indication that the probe was new, when in fact it was in the set (Cowan, Blume, & Saults, 2013; cf. Pashler, 1988).

Now suppose that we have four k estimates in different conditions: k_{s1} , the sounds-alone attention condition, k_{s2} , the sound test when both modalities were attended, k_{o1} , a colored-objects-alone condition, and k_{o2} , the test of a colored object when both modalities were attended. If there were no dual-task cost, the total number of items that could be held in the dual-task condition would be $k_{s1}+k_{o1}$. In reality, though, the total that could be held in this condition is only $k_{s2}+k_{o2}$. Therefore, the dual-task attention cost A can be expressed as

$$A = (k_{s1} + k_{o1}) - (k_{s2} + k_{o2})$$

This dual-task cost is equal to the estimated total number of items that can be held using central attentional mechanisms because when two modalities must be retained in working memory, on average each one loses $A/2$ compared to the unimodal situation.

Next, we can define modality-specific storage capabilities for sound items, S , and differently colored array objects, O . Performance is said to consist of what was remembered in the modality-specific way and what was remembered in the focus of attention, so that

$$k_{s1} = S + A \text{ and } k_{o1} = O + A$$

From the estimates of capacity in the unimodal conditions, k_{s1} and k_{o1} , along with the prior estimation of A , one can estimate S and O .

Using circles to represent memory in the unimodal verbal and colored objects conditions, a Venn diagram can be constructed to represent the dual-task cost in terms of an overlap between the circles, with the total area of the overlapped circles representing the total performance in the bimodal attention condition (Figure 1). On this Venn diagram, code-specific sound memory (S), code-specific colored-objects memory (O), and attention-based memory (A) are shown, each representing a number of items in working memory.

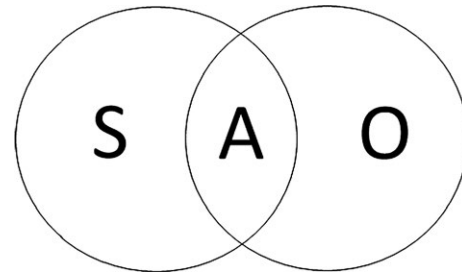


FIGURE 1 Theoretical division of the number of items remembered into three separate portions: those remembered from acoustic (sound) lists regardless of the attention condition (S), those remembered from visual object arrays regardless of the attention condition (O), and those remembered using an allocation of attention that does depend on the attention condition because of a common resource such as attention (A). In unimodal verbal (or visual object) trials the contents of Area A are entirely verbal (or visual), whereas in bimodal trials, Area A can be split between verbal list items and visual objects

In terms of this model, above, one attention-related hypothesis regarding the developmental increase of working memory capacity is that A increases with age, for an overall gain because of an expanding focus of attention; whereas it is alternatively possible that, with age, A decreases or remains unchanged but that both S and O increase, again for an overall gain. Thus, contrasting interpretations lead to the expectation of increased versus decreased dual-task conflict across age groups from the early elementary school years to adulthood.

1.4 | Limits of the approach

Note that the present approach can only *constrain* theories of working memory development in childhood, rather than completely determining which theory is correct. Each theoretical component has multiple possible interpretations. Attention-based memory (Component A) can refer to any process that can be reallocated from one stimulus set in working memory to another: either holding material from two domains in a general focus of attention concurrently (Cowan, 1988; Saults & Cowan, 2007) or dividing some mnemonic resource between domains, such as the attention-based refreshment of information that is otherwise held passively and would decay and become unavailable without this refreshment (Barrouillet et al., 2009; Vergauwe, Camos, & Barrouillet, 2014). The A component could increase with development if any of several mechanisms operating across domains improved, including the capacity of the focus of attention or the speed or efficiency of attention-based refreshing of information that happened to be passively held but could be forgotten without refreshment. Similarly, the code-specific components S and O could reflect either separate memory modules (Baddeley & Logie, 1999) or separate types of information in the activated portion of long-term memory, not separate modules but susceptible to the general principle of interference between items with similar features (Cowan, 1988). Moreover, developmental

increases in these code-specific components could come about in one of several ways: either because the knowledge underlying the representations in some way improves with age (Chi, 1978; but see Cowan, Ricker, Clark, Hinrichs, & Glass, 2015) or because with age, participants improve in the ability to off-load information from the focus of attention to more peripheral, less attention-demanding forms of storage for which different codes interfere with each other less (Cowan et al., 2014). In particular, it would presumably be less attention-demanding to divide refreshing between two stimulus sets that have been off-loaded to memory that persists for a while even without attention, compared to holding information from two modalities in the focus of attention without any benefit from automatic maintenance mechanisms. Ironically, this off-loading process could be a sophisticated use of attention that would ease the overall attentional demands on the participant and could, in terms of the model, lower the magnitude of the A component rather than raising it, or could allow more information to be stored without increasing the A component (Cowan et al., 2014), instead increasing S and O. This is potentially the case because the trade-off between

refreshing two stimulus sets in memory presumably is less than the trade-off between continual storage of both stimulus sets in the focus of attention. Given the issues that the present method cannot resolve regarding the interpretation of A, S, and O components, much of the General Discussion section will be dedicated to reviewing the possibilities that remain, given the important constraints emanating from the present results.

2 | EXPERIMENT 1

2.1 | Methods

The program that presented stimuli and collected data is available in the Open Science Framework at https://osf.io/gc6wb/?view_only=a102f57d7a1c4ad69ba01aeb64250c1c, including a separate file with the acoustic stimuli. Methods used in this experiment (see Figure 2) closely followed Experiment 1a of Cowan et al. (2014), with a few exceptions explained below. Arrays of colored squares were combined with lists of spoken digits and the conditions included attention to

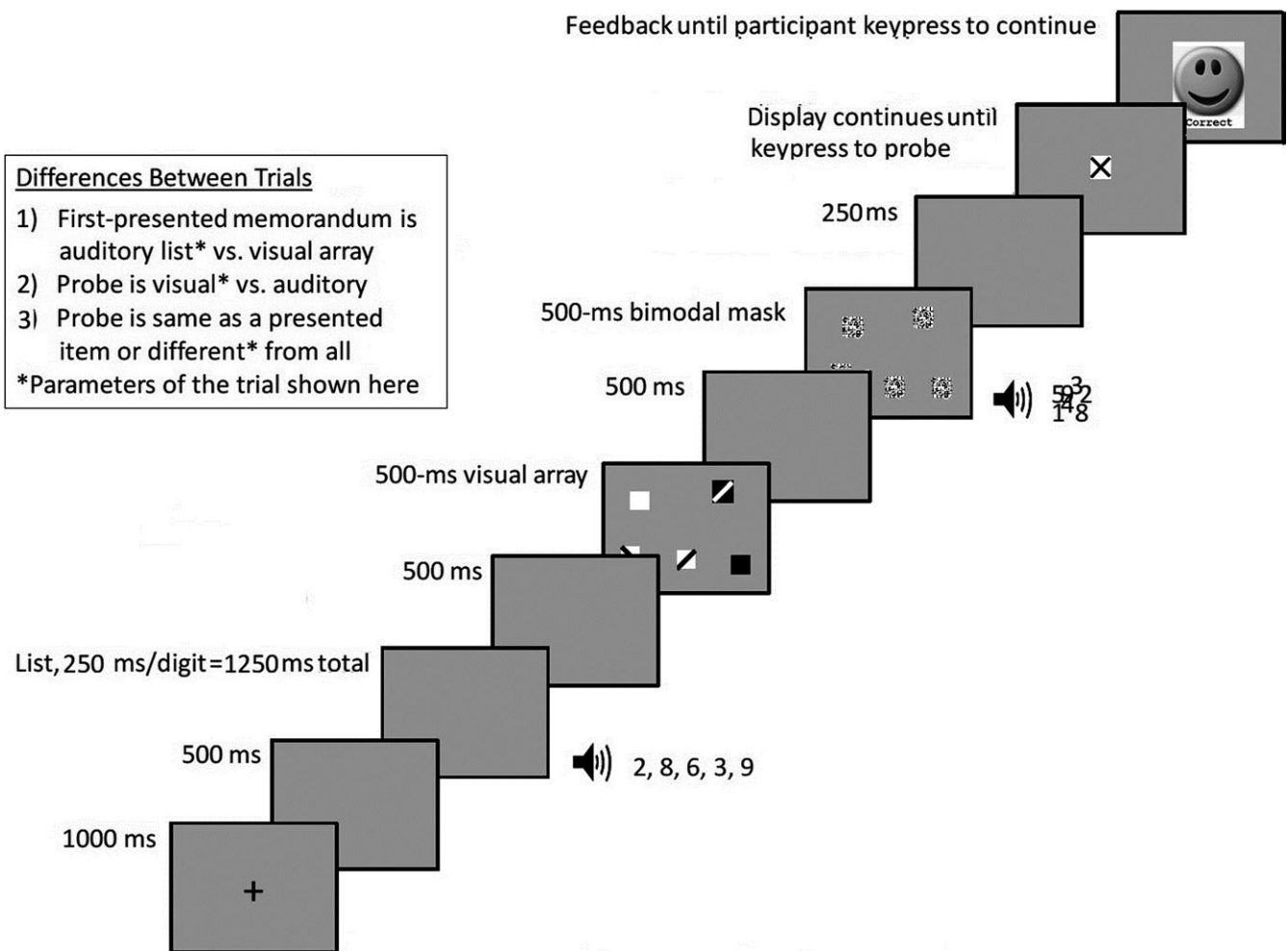


FIGURE 2 A detailed illustration of a trial in Experiment 1. The insert explains the different types of trials that occurred in three blocked memory load conditions. A *visual-load* block included only trials with visual probes; an *auditory-load* block included only trials with auditory probes; and a *bimodal-load* block included randomly intermixed trials with auditory and visual probes. Patterns of the visual array objects represent colors

visual stimuli only, acoustic stimuli only, or both modalities. Masks were presented in order to eliminate any residual sensory memory (Saults & Cowan, 2007). Then a probe was presented in an attended modality, a single spoken digit or a colored square in the center of the screen. The item was always either identical to an array or list item or different from all of the studied items.

The most important differences from Cowan et al. (2014, Experiment 1a) were (1) that we did not impose articulatory suppression or tapping as Cowan et al. did, an important simplification for children, and one that seems justified inasmuch as no effect of suppression was obtained by Cowan et al.; and (2) that the visual array probe was placed in the center of the screen, removing the spatial location cue to achieve more equivalence with the auditory list probe, for which there was no serial location cue.

2.1.1 | Participants

The participants with complete data include 31 children in the youngest age group, in Grades 1 and 2 and ages 6–8 (13 female and 18 male, $M = 7.68$ years, $SD = 0.52$); 30 in the second age group in in Grades 3 and 4 and ages 8–10 (20 female and 10 male, $M = 9.48$ years, $SD = 0.69$); 30 in the third age group, in Grades 5 through 7 and ages 10–13 (13 female and 17 male, $M = 11.81$ years, $SD = 0.94$); and 31 parents of the children (26 female and 5 male; ages 30.3 to 57.9 years, $M = 41.7$ years, $SD = 5.95$). These adults had an average of 17.27 years of education ($SD=2.59$), although one adult failed to report education. One child in the youngest age group was omitted because of an equipment malfunction. Non-age-normed mean Raven Progressive Matrices scores for the four age groups, respectively, were 27.26 ($SD = 8.50$), 34.30 ($SD = 7.96$), 40.97 ($SD = 7.27$), and 50.00 ($SD = 5.29$). According to traditional norms collected in 1979 (Raven, 2000), most of these means were around the 50th percentile, although the youngest age group mean was closer to the 75th percentile, possibly because younger children now have access to more visual games and puzzles contributing to skills underlying the Ravens test than they did in 1979. Newer norms (Pind, Gunnarsdóttir, & Jóhannesson, 2003) show all age group means to be close to the 50th percentile. No participant or guardian indicated that a participant had a developmental disability.

2.1.2 | Apparatus and stimuli

Auditory stimuli were the spoken digits 1–9, digitally recorded with 16-bit resolution at a sample rate of 22,500, spoken by an adult male voice. The recorded digits were temporally compressed to a maximum duration of 250 ms and presented at a pace of four digits per second. Temporal compression, to 65% to 95% of each word's original duration, without altering pitch was accomplished using the software Praat (Boersma & Weenink, 2009). The auditory mask combined all digits with onsets aligned. Digits and mask were presented to each ear with an intensity of 65–75 dB(A) using Audio-Technica ATH-M50WH Studio Monitor Headphones.

Visual stimuli were presented on a 17-inch cathode ray tube monitor (1024 by 768 pixels). Visual study arrays consisted of squares

whose colors were sampled without replacement from nine colors (*black, white, red, blue, green, yellow, orange, cyan, and magenta*). Study array squares were randomly positioned on the screen as described previously (Cowan et al., 2005) and were presented for 500 ms. Patterned masks consisted of multicolored squares of the same size in the same locations as the items in the studied array, also presented for 500 ms.

2.1.3 | Procedure

Each participant was tested individually in a quiet room. An experimenter was present throughout each session for the children. For adults, an experimenter was in the booth throughout the instructions and the practice of the first block of trials. Visual, auditory, and combined bimodal memory capacities were tested using a general procedure similar to Experiment 1a of Cowan et al. (2014) with the exceptions noted above (that here we did not use articulatory suppression, and that we used a central visual probe) and one other change. Without a suppression manipulation, we were able to use a simpler design with five counterbalanced blocks of first two unimodal memory blocks (20 visual-probe trials and 20 auditory-probe trials), then a bimodal memory block consisting of 80 intermixed visual-probe and auditory-probe trials, and finally two more unimodal memory blocks (20 visual-probe trials and 20 auditory-probe trials). The order of the unimodal conditions was further counterbalanced within and between participants, so that each participant performed the two unimodal tasks (auditory-probe and visual-probe conditions) in one order during the first two blocks and the reverse order in the last two blocks. Half of the participants in each age group performed the auditory-probe task first and the remaining participants in each age group performed the visual-probe task first.

An illustration of a trial is shown in Figure 2. All stimuli were presented on a screen with a uniform medium gray background. Each trial began with fixation cross presented in the center of the screen for 1000 ms, followed by a 500-ms blank screen. This was followed by an array of five different colored squares lasting 500 ms, and a sequence of five different digits presented four per second over headphones. The two modalities were separated by an interstimulus interval (ISI) of 500 ms; one modality was completed before the other one started and each modality occurred first on half of the trials. A blank screen appeared for 500 ms after the study items. Then five multicolored-mask squares appeared in the same locations as the squares in the study array. At the same time, an auditory mask, consisting of the combined digits, was presented via headphones. Thus, even though the targets were presented consecutively by modality (all auditory before all visual or vice versa), the masks were presented for both modalities at once. Although this caused the target-mask interval to vary between modalities, this arrangement was necessary to avoid distracting switches of attention that would result if the mask in each modality occurred at different times. In any case, target-mask intervals, along with modality presentation order, were counterbalanced across trials.

After the 500-ms mask, there was a 250-ms delay and then a probe was presented. The auditory probe was a spoken digit identical

to one in the study list or different from any in the study list. When an auditory probe occurred, a “?” appeared in the center of the screen. The visual probe was a square in the center of the screen that was the same color as a study square or a color different from any square in the study array. The participant was to press the “S” key if the probe was the same as a study item and the “D” key if the probe was different from all of the study items. The “?” or probe square remained on the screen until a response was recorded. Then participants saw an iconic picture of a smiling face to indicate a correct response or a frowning face to indicate an incorrect response. This feedback was displayed on the center of the screen until they pressed the spacebar to begin the next trial.

Each visual-load block consisted of eight practice and 20 experimental trials, in which participants were instructed to remember only the colored squares and were always tested with a probe square. Analogously, each auditory-load block consisted of eight practice and 20 experimental trials, in which participants were instructed to remember only the spoken digits and were always tested with a probe digit. Finally, the bimodal-load block consisted of 16 practice and 80 experimental trials, in which participants were instructed to remember both the colored squares and the spoken digits. In these blocks there were an equal numbers of auditory- and visual-probe trials randomly intermixed. In every block of trials, half of the trials for each probe modality and presentation order (visual-first or auditory-first) were change (different) trials. In experimental blocks with auditory probes, the probe digit in no-change (same) trials occurred in each serial position of the study list equally often. Overall, there were 160 experimental trials consisting of 80 unimodal memory load trials and 80 bimodal memory load trials, including equal numbers of trials with visual and auditory probes and with same and different as the correct response.

Any of the capacity estimates for an individual underlying the peripheral verbal, peripheral visual, or central working memory model components could go below zero given measurement error but, when this occurred, we adjusted such values to zero, the lowest theoretically meaningful value. In the four age groups, respectively, this adjustment occurred in 0.12, .05, .03, and .01 of all trials and resulted in a mean overestimate of capacity of 0.29, 0.08, 0.04, and 0.01 out of 5.00 items. Thus, the results as reported slightly underestimate the increases in capacity with age.

2.2 | Results and discussion

Proportion correct by condition, collapsed across the order of presentation of stimuli, and estimated items in working memory are shown in Table 1. The outcome of the derived analysis of visual, auditory/verbal, and central/attentional components in each age group is shown in Figure 3. The figure shows that the peripheral visual and acoustic/verbal, dedicated components of memory increased across age groups, whereas the central/attentional component did not increase. This pattern is consistent with the hypothesis that, as people mature, they become better able to encode more information in such a way that the representations for verbal/acoustic information do not interfere with those for visual object information, or vice versa.

In support of the pattern shown in Figure 3, an ANOVA was conducted with age group between participants and with two within-participant factors: the memory component (dedicated speech, dedicated visual object, and central attention) and the stimulus order within a trial (verbal-first or visual-first). All three variables produced main effects: age group, $F(3, 118) = 32.77, p = .000, \eta_p^2 = .45$; memory component, $F(2, 236) = 66.54, p = .000, \eta_p^2 = .36$; and stimulus order within a trial, $F(1, 118) = 25.60, p = .000, \eta_p^2 = .18$. The first two main effects can be seen in Figure 3: the effect of age, due primarily to improvements across age in the peripheral acoustic and peripheral visual parameters of the model, and the effect of model component, with central < visual < verbal component mean values. In Figure 3, one can also see the basis of a crucial interaction between age group and the memory component, $F(6, 236) = 7.43, p = .000, \eta_p^2 = .16$. The figure shows that the three components have very different developmental trajectories, with the verbal component increasing with age quickly, the visual component increasing more slowly, and the central component slightly decreasing. A separate analysis of the components showed an age group effect for the peripheral visual component, $F(3, 118) = 23.21, p = .000, \eta_p^2 = .37$, and for the peripheral verbal component, $F(3, 118) = 11.30, p = .000, \eta_p^2 = .22$, but not for the central component, $F(3, 118) = 1.72, p = .166, \eta_p^2 = .04$.

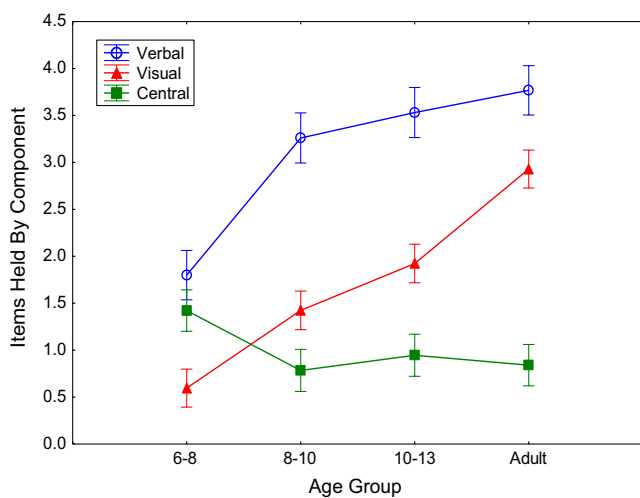
The one remaining effect was the interaction of the memory component with the stimulus order, that is, the presentation of auditory or visual memoranda first, $F(2, 236) = 6.29, p = .002, \eta_p^2 = .05$. It did not interact with age group. The mean verbal peripheral capacity (with SEM) was similar no matter whether the trial was verbal-first, 3.23 (0.16), or visual-first, 2.95 (0.15). The same was true for the mean central capacity: verbal-first, 0.85 (0.15); visual-first, 1.15 (0.14). However, visual peripheral capacity was considerably higher in the verbal-first condition, 2.11 (0.15), than in the visual-first condition, 1.33 (0.12). The difference could reflect either the loss of visual information over the longer delay for the first-presented modality, or the effect of verbal interference on visual performance. The overall capacity averaged 3.24 items for the auditory-first situation compared to 2.85 items for the visual-first, explaining the main effect of stimulus order across components reported above. Overall, these results suggest that the development of working memory has more to do with the dedicated visual and verbal components than it does with the central component.

There was also a superiority of verbal performance and special difficulty with visual information when it is followed by verbal information that also has to be retained. This same asymmetry between modalities has been reported in the adult literature (Morey & Bieler, 2013; Morey, Morey, van der Reijden, & Holweg, 2013).

The asymmetry between modalities suggests one possible explanation of the pattern of components and their changes across age groups. Specifically, it is possible that with development, the increasing levels of expertise on the number stimuli allowed chunking on the basis of knowledge (Miller, 1956) or allowed grouping (Cowan, Saults, Elliott, & Moreno, 2002; Farrell, 2012) in ways that protected the verbal information and thereby prevented mutual interference between verbal and visual information. As an example, if the 5-digit list (e.g.,

TABLE 1 Proportion correct and estimated items Mean (and SD) in working memory by condition in Experiment 1

Probe	Attention	Answer	1st presented	6–8 years	8–10 years	10–13 years	Adults
Proportion correct							
Auditory	Unimodal	Different	Auditory	.82 (.20)	.91 (.12)	.94 (.07)	.95 (.11)
Auditory	Unimodal	Same	Auditory	.73 (.20)	.84 (.12)	.9 (.11)	.94 (.08)
Auditory	Bimodal	Different	Auditory	.83 (.18)	.86 (.13)	.92 (.1)	.93 (.16)
Auditory	Bimodal	Same	Auditory	.69 (.22)	.80 (.14)	.87 (.14)	.91 (.14)
Auditory	Unimodal	Different	Visual	.84 (.18)	.92 (.12)	.97 (.06)	.94 (.11)
Auditory	Unimodal	Same	Visual	.74 (.21)	.83 (.14)	.91 (.10)	.93 (.11)
Auditory	Bimodal	Different	Visual	.76 (.23)	.93 (.11)	.92 (.11)	.91 (.14)
Auditory	Bimodal	Same	Visual	.66 (.20)	.78 (.14)	.86 (.12)	.86 (.19)
Visual	Unimodal	Different	Auditory	.72 (.16)	.82 (.17)	.80 (.14)	.89 (.12)
Visual	Unimodal	Same	Auditory	.60 (.19)	.64 (.22)	.71 (.18)	.79 (.15)
Visual	Bimodal	Different	Auditory	.65 (.19)	.78 (.18)	.79 (.16)	.77 (.16)
Visual	Bimodal	Same	Auditory	.50 (.22)	.52 (.22)	.60 (.19)	.79 (.16)
Visual	Unimodal	Different	Visual	.68 (.18)	.75 (.18)	.80 (.16)	.87 (.13)
Visual	Unimodal	Same	Visual	.54 (.19)	.49 (.22)	.62 (.22)	.81 (.13)
Visual	Bimodal	Different	Visual	.65 (.22)	.73 (.13)	.75 (.17)	.80 (.17)
Visual	Bimodal	Same	Visual	.41 (.18)	.48 (.18)	.53 (.16)	.70 (.13)
Estimated items in working memory							
Auditory	Unimodal	n/a	Auditory	3.14 (1.59)	4.11 (0.72)	4.43 (0.6)	4.65 (0.56)
Auditory	Bimodal	n/a	Auditory	3.03 (1.52)	3.80 (0.87)	4.26 (0.81)	4.39 (1.09)
Auditory	Unimodal	n/a	Visual	3.30 (1.59)	3.98 (0.90)	4.52 (0.55)	4.57 (0.71)
Auditory	Bimodal	n/a	Visual	2.63 (1.52)	3.73 (1.03)	4.22 (0.70)	4.21 (1.20)
Visual	Unimodal	n/a	Auditory	2.28 (1.21)	2.72 (1.30)	3.12 (1.14)	3.71 (1.11)
Visual	Bimodal	n/a	Auditory	1.32 (1.27)	1.99 (1.25)	2.47 (1.17)	3.50 (1.33)
Visual	Unimodal	n/a	Visual	1.75 (1.27)	1.70 (1.42)	2.62 (1.34)	3.83 (0.91)
Visual	Bimodal	n/a	Visual	0.65 (0.85)	1.41 (1.23)	1.85 (1.05)	2.98 (1.06)

**FIGURE 3** In Experiment 1, as a function of the age group, performance parameters for verbal list dedicated memory (open circles; corresponding to S in Figure 1), visual object array dedicated memory (triangles; corresponding to O in Figure 1), and flexible central memory (squares; corresponding to A in Figure 1). Error bars are standard errors

83541) on a bimodal trial happened to correspond to two known sequences (e.g., 835 as a known house number and 41 as one's own age), a two-chunk recoding of the verbal information would reduce the need for attention to help retain that information, which in turn would allow the devotion of attention primarily to the visual stimuli. Experiment 2 addresses this possible account of the data using stimuli for which verbal knowledge could not play such a role, and in which the asymmetry between modalities was largely eliminated.

3 | EXPERIMENT 2

In this second experiment, we hoped to determine whether the asymmetry between visual and acoustic performance is necessary in order to observe the pattern of results in which the peripheral, but not the central, component increased with age. Toward this end, we replaced the spoken digits of Experiment 1 with special tonal stimuli, knowing that visual arrays of colors and tones do compete for attention (Morey, Cowan, Morey, & Rouder, 2011; Stevanovski & Jolicoeur, 2007). Our tones were developed for a previous adult study (Li, Cowan, & Saults,

2013, Experiment 3). The stimuli are tones produced by a synthesizer, in different instruments to decrease the task difficulty. In order to reduce knowledge, they occur at pitches that increase in constant ratio steps but do not fit any known musical scale. We expected that these stimuli would decrease the advantage of acoustic stimuli over visual stimuli by greatly decreasing any role of knowledge compared to Experiment 1. The purpose of the second experiment was to determine whether, under these modified conditions, the peripheral components would still increase and the central component would still not increase across development.

3.1 | Method

The program that presented stimuli and collected data is available in the Open Science Framework at https://osf.io/gc6wb/?view_only=a102f57d7a1c4ad69ba01aeb64250c1c, including a separate file with the acoustic stimuli.

3.1.1 | Participants

Our participant sample was quite similar to that of Experiment 1, with 31 participants in Group 1 (16 female, 16 male; $M = 7.73$ years, $SD = 0.78$), 32 in Group 2 (14 female, 18 male; $M = 9.83$ years, $SD = 0.63$), 30 in Group 3 (15 female, 15 male; $M = 12.19$ years, $SD = 0.98$), and 32 parents in Group 4 (27 female, 5 male; $M = 41.32$ years, $SD = 5.92$). The age of one such adult was recorded only in years. The adults had a mean of 16.66 years of education, $SD = 2.55$. The raw Ravens Progressive Matrices scores were similar to Experiment 1 as well: in the four age groups, respectively, $M = 33.50$, $SD = 6.70$; $M = 40.66$, $SD = 7.16$; $M = 45.80$, $SD = 6.13$; and $M = 50.94$, $SD = 5.97$. All groups were closest to the 75th percentile according to the norms of Raven (2000) and between the 50th and 75th percentiles according to the newer norms of Pind et al. (2003). The Ravens test was not completed by one of the adults. No participant or guardian indicated that the child had a developmental disability.

3.1.2 | Apparatus, stimuli, and procedure

The set-up was identical to Experiment 1 except that the aforementioned tonal stimuli were used in place of spoken digits. We used the lowest 9 of 12 easily discriminable tones developed by Li et al. (2013, Experiment 3). Li et al. used instruments generated with GarageBand (Apple Inc., Cupertino, CA), a program in the Macintosh Operating System. In the stimulus set, each fundamental frequency was associated with a different instrument to provide redundant cues to sound identity. We used *Trumpet Section* at 200 Hz; *Smooth Clav* at 262 Hz, *Classic Rock Organ* at 343 Hz, *Negril Bass* at 450 Hz, *Tenor Sax* at 589 Hz, *Space Harpsichord* at 772 Hz, *Grand Piano* at 1011 Hz, *Live Pop Horns* at 1324 Hz, and *Aurora Bell* at 1735 Hz. There was a 31% frequency difference between each two adjacent tones. Each sound was 500 ms long in Li et al. but, for the present study, was truncated to 125 ms. The multi-tone auditory mask was 500 ms long, consisting of a series of each of the 9 tones for 9/500 ms, connected with no gap

and presented along with the visual mask as in Experiment 1. There was a brief learning exposure to the continuum of tones.

The rationale for the selection of tone frequencies cannot be rephrased much without risking the loss of important information, so we quote it here from Li et al. (2013, p. 150):

We wanted the pitches of our 12 tones to be as far apart as possible, so they would be easy to discriminate, but still within a range with similar difference limens for frequency change, which increases sharply beyond 4000 Hz (Sek & Moore, 1995). We also wanted them to differ from familiar musical notes. Thus, our lowest tone was about 35 cents above the G below middle C (G3), while our highest tone was about 23 cents below B7, the second highest note on an 88-key piano (100 cents = 1 semitone). A 31% difference between tones avoids familiar musical intervals and harmonic relationships between tones. Adjacent semitones in music differ by about 5.9% (precisely $2^{1/12}$) in 12-tone equal temperament, the common tuning system for Western music (Burns & Ward, 1999). Although our stimuli spanned about four octaves, no tone in our set had a simple harmonic relationship with another tone. For example, the second harmonic of 200 Hz is 800 Hz, but the closest frequency to that in our set was 771.6 Hz. Avoiding octaves minimizes the tendency to confuse two tones with different pitch height but equal chroma on the basis of octave generalization (Shepard, 1982).

There were slight changes in the stimulus timing compared to Experiment 1. Instead of series of digits lasting up to 250 ms, we presented tones at the same pace, but with a presentation schedule of 125 ms on, 125 ms off, within that pace so that the stimuli would not seem to blend together. (Such blending did not occur in Experiment 1 because of the speech waveform envelopes.) In Experiment 1 there was an additional 500-ms delay between the end of the second stimulus set and the onset of the mask, and a 250-ms delay between the mask offset and probe onset (Figure 2); but in Experiment 2 we reversed these delay lengths, with 250 ms between stimuli and mask and with 500 ms between mask and probe. This change was made to accommodate the different lengths of the digits versus tones; thus, in the tones-second order, the silent time from the end of the last tone to the onset of the mask was 125+250 ms, which made the timing sound similar to that of Experiment 1.

Below-zero capacities were adjusted to zero and in the four age groups, respectively, this adjustment occurred in 0.20, .10, .12, and .02 of all trials and resulted in a mean overestimate of capacity of 0.48, 0.33, 0.56, and 0.15 out of 5.00 items. Thus, as in Experiment 1, the results as reported slightly underestimate the increases in capacity with age.

3.2 | Results and discussion

The mean proportions correct and estimated items in working memory for each condition in each age group can be found in Table 2,

collapsed across stimulus order. For the derived central and peripheral components, Figure 4 shows both similarities and important differences from Experiment 1. Like that experiment, both the visual and acoustic dedicated memory parameters increased markedly across age groups. Also like that experiment, the central attentional component failed to increase across age groups.

Unlike Experiment 1, the present experiment with colors and tones shows rather comparable levels for the visual and auditory dedicated components. Thus, in contrast to Experiment 1 with its verbal stimuli, the pattern of performance (with increasing peripheral components but not an increasing central component with development) cannot be explained by any sort of asymmetry between the modalities that might remove one modality from the need for attention.

Support for these statements comes from an ANOVA with the same factors as in Experiment 1. There was a main effect of age group, $F(3, 122) = 24.19, p = .000, \eta_p^2 = .37$, and of the memory parameter, $F(2, 244) = 23.35, p = .000, \eta_p^2 = .16$. Most importantly, there was an interaction of the age group with the memory parameter, $F(6, 244) = 5.75, p = .000, \eta_p^2 = .13$. As in Experiment 1, separate analyses indicated an age group effect for the peripheral visual component, $F(3,$

$122) = 15.14, p = .000, \eta_p^2 = .27$, and the peripheral acoustic component, $F(3, 122) = 8.64, p = .000, \eta_p^2 = .18$, but not the central component, $F(3, 122) = 2.05, p = .111, \eta_p^2 = .05$.

Unlike Experiment 1, no effects involving stimulus order approached significance, $p > .24$ for each effect. In Experiment 1, participants had special difficulty with the visual stimuli when they were presented first on a trial, and it was unclear whether that difficulty stemmed from the passage of time or from interference caused by the subsequent verbal stimuli. Given that no such effect was obtained in Experiment 2, it was apparently interference from the verbal stimuli that was critical, in line with the possibility that some processes involving verbalization (such as color categorization) could play a role in maintenance of the visual stimuli, even though verbal rehearsal per se does not seem to play such a role (Morey & Cowan, 2004).

The finding that there was very little central component for the tones in adults is surprising, in that it conflicts with two other studies of attention-sharing between colors and tones in working memory (Morey et al., 2011; Stevanovski & Jolicoeur, 2007). One important departure from these previous studies is that our tones had different timbres to make the stimuli easier for children. Adults may be able to use this timbre information to help eliminate the need for central attention

TABLE 2 Proportion correct and estimated items Mean (and SD) in working memory by condition in Experiment 2

Probe	Attention	Answer	1st presented	6–8 years	8–10 years	10–13 years	Adults
Proportion correct							
Auditory	Unimodal	Different	Auditory	.61 (0.19)	.63 (0.22)	.66 (0.24)	.63 (0.16)
Auditory	Unimodal	Same	Auditory	.53 (0.25)	.66 (0.18)	.62 (0.17)	.76 (0.18)
Auditory	Bimodal	Different	Auditory	.58 (0.22)	.58 (0.20)	.60 (0.22)	.59 (0.20)
Auditory	Bimodal	Same	Auditory	.62 (0.20)	.64 (0.19)	.67 (0.18)	.79 (0.17)
Auditory	Unimodal	Different	Visual	.58 (0.18)	.62 (0.20)	.62 (0.21)	.58 (0.23)
Auditory	Unimodal	Same	Visual	.64 (0.20)	.68 (0.16)	.71 (0.18)	.79 (0.19)
Auditory	Bimodal	Different	Visual	.59 (0.25)	.60 (0.23)	.54 (0.22)	.56 (0.19)
Auditory	Bimodal	Same	Visual	.61 (0.26)	.68 (0.16)	.72 (0.18)	.88 (0.14)
Visual	Unimodal	Different	Auditory	.70 (0.19)	.73 (0.20)	.78 (0.21)	.83 (0.18)
Visual	Unimodal	Same	Auditory	.56 (0.17)	.63 (0.20)	.69 (0.16)	.78 (0.15)
Visual	Bimodal	Different	Auditory	.67 (0.20)	.80 (0.18)	.75 (0.21)	.80 (0.15)
Visual	Bimodal	Same	Auditory	.49 (0.17)	.53 (0.17)	.60 (0.16)	.73 (0.18)
Visual	Unimodal	Different	Visual	.72 (0.18)	.73 (0.20)	.74 (0.19)	.82 (0.13)
Visual	Unimodal	Same	Visual	.52 (0.19)	.64 (0.21)	.62 (0.17)	.78 (0.19)
Visual	Bimodal	Different	Visual	.68 (0.20)	.68 (0.19)	.70 (0.20)	.78 (0.20)
Visual	Bimodal	Same	Visual	.45 (0.19)	.45 (0.20)	.55 (0.19)	.70 (0.15)
Estimated items in working memory							
Auditory	Unimodal	n/a	Auditory	1.49 (1.59)	2.21 (1.40)	2.09 (1.43)	3.08 (1.39)
Auditory	Bimodal	n/a	Auditory	1.65 (1.28)	1.89 (1.43)	2.27 (1.50)	3.11 (1.50)
Auditory	Unimodal	n/a	Visual	1.97 (1.68)	2.43 (1.26)	2.68 (1.48)	3.10 (1.76)
Auditory	Bimodal	n/a	Visual	1.78 (1.79)	2.14 (1.47)	2.40 (1.60)	3.80 (1.37)
Visual	Unimodal	n/a	Auditory	1.84 (1.23)	2.28 (1.56)	2.93 (1.16)	3.58 (1.10)
Visual	Bimodal	n/a	Auditory	1.18 (1.25)	1.99 (1.07)	2.30 (1.28)	3.20 (1.25)
Visual	Unimodal	n/a	Visual	1.65 (1.49)	2.52 (1.49)	2.40 (1.36)	3.57 (1.34)
Visual	Bimodal	n/a	Visual	1.22 (1.14)	1.29 (1.05)	1.79 (1.37)	2.95 (1.18)

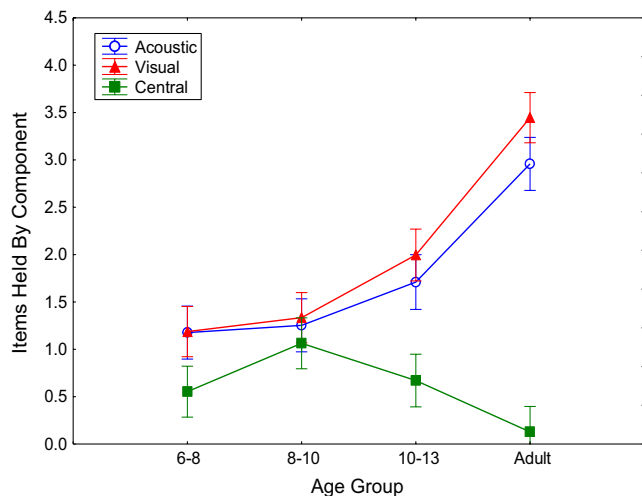


FIGURE 4 In Experiment 2, as a function of the age group, performance parameters for acoustic (tone) list dedicated memory (open circles; corresponding to S in Figure 1), visual object array dedicated memory (triangles; corresponding to O in Figure 1), and flexible central memory (squares; corresponding to A in Figure 1). Error bars are standard errors

in maintaining what they can from the sound sequences, once a mental representation is set up.

4 | GENERAL DISCUSSION

Two experiments have shown that between the middle elementary school years and adulthood, children improve in their ability to retain two very different sets of information in working memory for an upcoming item recognition task in such a way that the interference between them, measured in items, is similar across development despite increasing visual and acoustic memory levels with development. This developmental pattern occurred no matter whether the items were color arrays and digit lists (as in Experiment 1) or color arrays and lists of mistuned musical sounds (as in Experiment 2).

The benefit of the present results is intricately tied to the theoretical analysis, which produces separate estimates of the number of items held in a type or part of storage dedicated to visual information given our procedure; a type or part dedicated to acoustic or verbal information; and a type or part that is flexible, sacrificing some of the items in one modality as necessary in order to accommodate some items in the other modality. Within that analysis, one cannot say that the results that were obtained were predictable, inasmuch as different ways of thinking about the situation lead to different intuitions and predictions. There is fairly widespread agreement that attention capabilities increase with age in childhood (Ristic & Enns, 2015). That being the case, it could be anticipated that the flexible, central, attention-demanding part of the system would increase with age. That is not what the data reveal. Clearly there are increases in attention capability, but they do not predominate in our data unless attention processes include the means of avoiding conflict between two different types of input.

One way of thinking about developmental change is that, as children get older, they find new ways to relieve the strain on attention, storing the information in attention-free ways, perhaps forming new groups of items that are memorized. That kind of process corresponds well to our results, which show more dedicated storage of each modality without more accompanying intermodal interference. This set of results may be related to the finding of Janczyk, Büschelberger, and Herbolt (2017) of a developmental decrease in cross-talk interference between two speeded tasks (decreasing effects of incompatibility when the two responses were to be made on opposite sides of the body).

Note that the greater efficiency of off-loading information to forms of storage that minimize interference between stimulus sets need not increase the A component, and in principle could decrease it. In combination with increases in the S and O components, our finding of a slightly decreasing A component across age groups is fully compatible with attention processes that improve across age groups. Participants may find that the most effective use of attention is not to store information in it but to use it in a transient manner to insulate the two stimulus sets from one another in working memory. The effect of attention on memory would then be amplified in more mature participants, allowing more storage overall; but conflict between stimulus sets would not be commensurately increased, potentially even diminishing the A component. This kind of proposal for how our results came about is in keeping with recent neuroscientific evidence of transient neural effects of attention with mnemonic benefits outlasting the neural activity (Wallis, Stokes, Cousijn, Woolrich, & Nobre, 2015), a process that could underlie attentional refreshing (e.g., Vergauwe et al., 2014).

If working memory development is metaphorically considered to be an elephant, with each new study like a blind man exploring the nature of a unique part of the elephant, the present results add an important new appendage for consideration; perhaps even the trunk. One of the key issues in working memory development has been to understand how the development of attention is related to the development of working memory. Here we show that, with development, the need to divide attention across modalities does not increase (at least when stimuli in the two modalities are presented at different times in the trial), whereas the ability to retain items from two very different sets without mutual interference between them increases. Future work can now be designed to ask questions about why this pattern occurs. Seven possibilities taken from the literature are as follows, with an initial assessment of each one.

4.1 | Possible mechanisms of the obtained developmental separation between modalities

4.1.1 | Deployment of attentional filtering?

Traditionally, a distinction has been made between types of attention that are exogenous, or controlled by external stimuli, versus endogenous, or controlled by the mind. The latter type of attention is the type that develops across the elementary school years according to many different procedures (for a review see Ristic & Enns, 2015). Closely related work discusses the development of executive function



(Davidson, Amso, Cruess Anderson, & Diamond, 2006). One potentially relevant application of attention and executive function, which has been put forward in adult studies as a basis of individual differences, is the ability to attend more to stimuli that are of greater task relevance (e.g., Vogel, McCollough, & Machizawa, 2005). In the present situation, the result presumably would be inappropriate attention by young children to irrelevant stimuli in the unimodal test blocks (cf. Cowan, Fristoe, Elliott, Brunner, & Sauls, 2006; Hagen, 1967; Plebanek & Sloutsky, 2017). Although we do not know if this occurred, in young children it would produce a diminished central component by decreasing unimodal task performance (for which the stimuli in the irrelevant modality were still presented) and thereby decreasing the difference between unimodal and bimodal trials. It would thus not explain the relatively constant central component across age groups. Moreover, this attention-allocation concept has failed to account for some other situations. For example, Cowan, Morey, AuBuchon, Zwillig, and Gilchrist (2010) and Cowan, AuBuchon, Gilchrist, Ricker, and Sauls (2011) found that even first-grade children displayed lower recall for the colors presented in a rarely-tested shape compared to colors presented in an often-tested shape (tested on 80% of the trials in the block), to the same extent as older participants, even though the young children performed more poorly.

4.1.2 | Development of a proactive stance?

Another developmental concept is that younger children react to each stimulus as it occurs rather than planning ahead, that is, they maintain a reactive rather than a proactive stance. For example, Morey, Mareva, Lelonkiewicz, and Chevalier (2017) examined eye movements during a sequence of visual objects to be remembered and showed that children in the late preschool and early school years did less to remember the sequence during its presentation, waiting until the end of the sequence and apparently trying to make up for that paucity of proactive, encoding-related activity during the retention interval. Applied to the present findings, one might expect that children would not adequately carry out mnemonic activities during the presentation of the first stimulus set, losing that set when a second set to be remembered was presented later in the trial. However, a proactive stance cannot provide a consistent account of our stimulus order effects. In Experiment 1 we found that participants had trouble specifically with visual information that was followed by verbal information in bimodal trials; perhaps some of these participants failed to take a proactive stance toward the first-presented visual items and therefore lost more of those items, but that stimulus order \times modality interaction did not interact with age. In Experiment 2, there was no such effect of stimulus order.

4.1.3 | Development of verbal rehearsal?

It is theoretically possible that as children get older, they use verbal rehearsal more often (Flavell et al., 1966; Ornstein & Naus, 1978). In our study, doing so would allow the rehearsed stimuli to make fewer demands on attention (Guttentag, 1984), which would reduce the interference between modalities. In Experiment 1, presumably the

digits would be rehearsed in this way. In Experiment 2, which involved non-rehearsable, non-musical notes played by different instruments, perhaps the colors would be rehearsed instead. Militating against this interpretation, though, Morey and Cowan (2004) found that articulatory suppression had no effect on memory for arrays of categorical colors such as this. Moreover, Cowan et al. (2011), who examined memory for slowly-presented (1/s) sequences of colors embodied in more- versus less-relevant shapes, obtained the same basic pattern of results as when the items were presented in arrays (Cowan et al., 2010) and obtained no change in the developmental pattern of results using articulatory suppression. It seems unlikely, therefore, that covert rehearsal can account for the developmental pattern in the present study, either. Elsewhere, the evidence for the role of rehearsal in development has been questioned (e.g., Jarrold & Citroën, 2013).

4.1.4 | Development of knowledge?

It is quite clear that as children develop, the increasing access to knowledge makes it easier to form larger, more meaningful chunks of information and thereby reduces the memory load (e.g., Chi, 1978). Applied to the present results, it seems theoretically possible that older individuals, having had more experience with colors, would be able to combine them in a way that would reduce the memory load compared to children. This presumably would reduce the amount of material to be held by the attention-based, flexible component of memory. This kind of theory was, however, tested by Cowan et al. (2015). They presented arrays of five English letters or three unfamiliar characters. Although there was a clear advantage for the letters, which grew with age, when the results were standardized within each type of material (but across age groups), the finding was that the developmental trend was nearly identical for both kinds of materials. Knowledge of highly familiar stimuli like English letters or categorical colors may reach an asymptotically high level by early in the elementary school years, so use of this knowledge seems unlikely to be the major reason why working memory performance improves for such materials across the elementary school years and beyond.

4.1.5 | Development of pattern detection?

Jiang, Chun, and Olson (2004) found that adults were able to find patterns within arrays that increased the number of items recalled. The patterns may not be just simple chunks, but rather statistical structure, or apparent patterns, that can be detected on particular trials even if the patterns emerge by chance (Brady & Tenenbaum, 2013). On one hand, it is possible that this pattern-detection does not develop, given that even infants can detect and use statistical structure (e.g., Saffran, Johnson, Aslin, & Newport, 1999). On the other hand, the human pattern-encoding mechanisms that have been proposed are quite complex, including hierarchical Bayesian modeling systems in which there is an abstract level at which it is not the patterns per se that are at issue, but rather the abstracted pattern of patterns that occur in the world (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Thus, it is quite possible, though unproven, that the pattern-recognition mechanism could become more sophisticated with age. Applied to the present results,

this could mean that older participants form better structures with which to encode and quickly memorize parts of stimulus arrays, reducing the memory load and the reliance on attention to hold individual objects. Translated into specifics, the developmental change compatible with this framework could occur for various subtle reasons. As just one example, it is possible that the older participants understand better how computer randomization works, and therefore are less likely to make the mistake of expecting each display to be somehow similar to the previous one. Knowledge, chunking, and pattern detection may work together in that better knowledge allows better recognition of possible patterns and multi-item chunks (Brady & Tenenbaum, 2013; Miller, 1956). A more advanced participant may realize that coincidental combinations can occur and may actively search for such patterns for mnemonic benefits. Three colors from an array might be encoded together as the colors of a particular national flag; successive numbers from a sequence might be encoded as a regular progression (e.g., 6, 4, 2); or a tone progression might accidentally be close enough to a known melody to be encoded as a distorted version of that melody.

4.1.6 | Development of the speed of processing?

It has been well documented throughout the years that the speed of various processes increases with childhood development (e.g., Barrouillet et al., 2009; Case et al., 1982; Kail & Park, 1994). By rehearsing or refreshing information in working memory more quickly, the theory goes, there is less decay of information from working memory; more items can be preserved until the memory test. This kind of theory can account for developmental increases in working memory for any kind of materials. It would predict, however, that forgetting would be more severe for the first set of items compared to the second, inasmuch as there is more time in which to lose more of those items. That pattern was obtained for visual-first materials in Experiment 1 (although not differentially by age), but not in Experiment 2. Although there may be a role for speed of processing, it is not clear that it could account for the development of the ability to retain information in two nonverbal domains, as in the second experiment, without an increase in the central component of the model with age.

The speed of processing has also been addressed in prior work. Although all of the present materials were presented at a fairly fast pace, Cowan et al. (2011) presented visual arrays in a slower, one item/s sequential manner to allow better encoding and found the same developmental pattern as Cowan et al. (2010) using rapid, concurrent arrays. Cowan et al. (2011) also found that articulatory suppression introduced to prevent rehearsal did not change the developmental pattern for these visual arrays.

4.1.7 | Development of a dynamic field?

Simmering and Miller (2016) account for capacity development through an increase in the precision of representations in working memory, in a dynamic field model in which there are excitations related to the representations of stimulus qualities and inhibitory interactions between similar representations. More precise

representations with age would reduce the inhibitory interactions within a stimulus set, but it is difficult to see how this process can account for a developmental difference in the concurrent retention of two very different stimulus sets (e.g., in Experiment 2, a color array and an instrument-based but non-musical tone sequence), with greater interference between sets when they are both task-relevant as compared to when one set is task-irrelevant. Although this model might well be expanded to account for the development of the ability to keep separate more material in the two sets using attention, it is not yet clear how that would occur.

Finally, in the context of dynamic systems, it is also worth remembering that the role of auditory sensory memory may be uncertain; it could still contribute something despite the use of a mask, as was suggested by Nees (2016). Cowan, Nugent, Elliott, and Saults (2000) found age differences in the temporal decay of auditory sensory memory derived from lists unattended at the time of their presentation and then post-cued for recall, but with no indication of whether these age difference can survive a mask.

4.2 | Conclusion

Using a relatively simple, capacity-based model of working memory in dual-task situations (Cowan et al., 2014), we have shown that the development of working memory for very different materials, presented one after another, includes development of the ability to remember more of both without a concomitant increase in interference between the two sets in working memory. Among the explanations we have considered, a promising one is development in the ability to detect patterns within the random arrangements of stimuli that were presented, reducing the memory load. There may also be a role for the development of attention and capacity in this finding. A memory load reduces the resources that remain for processing (Chen & Cowan, 2009; Vergauwe et al., 2014) and perhaps that may in turn decrease the resources left in young children to find a better encoding or detect patterns among the stimuli, which would have allowed separate maintenance of more visual and acoustic information. Indirectly supporting this possibility, Cowan et al. (2010) found that overloading working memory prevented young children from even allocating attention rationally, which they were able to do when the working memory load was smaller. The present work is the beginning of what we see as a productive endeavor to use intermodal interference to understand working memory development in childhood.

ACKNOWLEDGEMENTS

This research was supported by NIH Grant R01 HD-21338 to NC. The idea for Experiment 1 came from JSS.

REFERENCES

- Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 8 (pp. 47–89). New York: Academic Press.



- Baddeley, A.D., & Logie, R.H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge: Cambridge University Press.
- Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology, 45*, 477–490.
- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.2.02) [Computer program]. Retrieved from: <http://www.praat.org/>
- 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*, 85–109.
- Burns, E.M., & Ward, W.D. (1999). Intervals, scales, and tuning. In D. Deutsch & D. Deutsch (Eds.), *The psychology of music* (2nd ed., pp. 215–264). San Diego, CA: Academic Press.
- Case, R. (1995). Capacity based explanations of working memory growth: A brief history and reevaluation. In F.E. Weinert & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 23–44). Mahwah, NJ: Erlbaum.
- Case, R., Kurland, D.M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.
- Chen, Z., & Cowan, N. (2009). How verbal memory loads consume attention. *Memory and Cognition, 37*, 829–836.
- Chi, M.T.H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*, 163–191.
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.
- Cowan, N. (2005). *Working memory capacity*. Hove, East Sussex: Psychology Press. [Psychology Press and Routledge Classic Edition with new foreword, 2016]
- Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science, 11*, 239–264.
- Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin and Review, 24*, 1158–1170.
- Cowan, N., & Alloway, T. (2009). The development of working memory. In M. Courage & N. Cowan (Eds.), *The development of memory in infancy and childhood* (pp. 303–342). Hove, East Sussex: Psychology Press.
- 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.
- Cowan, N., Blume, C.L., & Saults, J.S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 731–747.
- 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 and Cognition, 34*, 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., Nugent, L.D., Elliott, E.M., & Saults, J.S. (2000). Persistence of memory for ignored lists of digits: Areas of developmental constancy and change. *Journal of Experimental Child Psychology, 76*, 151–172.
- Cowan, N., Ricker, T.J., Clark, K.M., Hinrichs, G.A., & Glass, B.A. (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science, 18*, 132–145.
- Cowan, N., Saults, J.S., & Blume, C.L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General, 143*, 1806–1836.
- Cowan, N., Saults, J.S., Elliott, E.M., & Moreno, M. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46*, 153–177.
- Davidson, M.C., Amso, D., Cruess Anderson, L., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.
- Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119*, 223–271.
- Flavell, J.H., Beach, D.H., & Chinsky, J.M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37*, 283–299.
- Gathercole, S.E., Pickering, S.J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177–190.
- Guttentag, R.E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology, 37*, 92–106.
- Hagen, J.W. (1967). The effect of distraction on selective attention. *Child Development, 38*, 685–694.
- Halford, G.S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11*, 236–242.
- Henry, L. (2012). *The development of working memory in children*. London: Sage.
- Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology, 47*, 72–87.
- Janczyk, M., Büschelberger, J., & Herbold, O. (2017). Larger between-task crosstalk in children than in adults: Behavioral results from the backward crosstalk paradigm and a diffusion model analysis. *Journal of Experimental Child Psychology, 155*, 95–112.
- Jarrold, C., & Citroën, R. (2013). Reevaluating key evidence for the development of rehearsal: Phonological similarity effects in children are subject to proportional scaling artifacts. *Developmental Psychology, 49*, 837–847.
- Jiang, Y., Chun, M.M., & Olson, I.R. (2004). Perceptual grouping in change detection. *Perception and Psychophysics, 66*, 446–453.
- Kail, R., & Park, Y.S. (1994). Processing time, articulation time, and memory span. *Journal of Experimental Child Psychology, 57*, 281–291.
- Li, D., Cowan, N., & Saults, J.S. (2013). Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception, and Psychophysics, 75*, 145–160.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.
- Morey, C.C., & Bieler, M. (2013). Visual short-term memory always requires attention. *Psychonomic Bulletin and Review, 20*, 163–170.
- Morey, C.C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin and Review, 11*, 296–301.
- Morey, C.C., Cowan, N., Morey, R.D., & Rouders, J.N. (2011). Flexible attention allocation to visual and auditory working memory tasks: Manipulating reward induces a tradeoff. *Attention, Perception, and Psychophysics, 73*, 458–472.
- Morey, C.C., Mareva, S., Lelonkiewicz, J., & Chevalier, N. (2017). Gaze-based rehearsal in children under 7: A developmental investigation of eye movements during a serial spatial memory task. *Developmental Science*, <https://doi.org/10.1111/desc.12559>
- Morey, C.C., Morey, R.D., van der Reijden, M., & Holweg, M. (2013). Asymmetric cross-domain interference between two working memory

- tasks: Implications for models of working memory. *Journal of Memory and Language*, 69, 324–348.
- Nees, M.A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology*, 7, 1892.
- Ornstein, P.A., & Naus, M.J. (1978). Rehearsal processes in children's memory. In P.A. Ornstein (Ed.), *Memory development in children* (pp. 69–99). Hillsdale, NJ: Erlbaum.
- Pascual-Leone, J., & Johnson, J. (2011). A developmental theory of mental attention: Its applications to measurement and task analysis. In P. Barrouillet & V. Gaillard (Eds.), *Cognitive development and working memory: A dialogue between neo-Piagetian theories and cognitive approaches* (pp. 13–46). Hove, East Sussex: Psychology Press.
- Pascual-Leone, J., & Smith, J. (1969). The encoding and decoding of symbols by children: A new experimental paradigm and a neo-Piagetian model. *Journal of Experimental Child Psychology*, 8, 328–355.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics*, 44, 369–378.
- Pind, J., Gunnarsdóttir, E.K., & Jóhannesson, H.S. (2003). Raven's Standard Progressive Matrices: New school age norms and a study of the test's validity. *Personality and Individual Differences*, 34, 375–386.
- Plebanek, D.J., & Sloutsky, V.M. (2017). Costs of selective attention: When children notice what adults miss. *Psychological Science*, 28, 723–732.
- Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, 41, 1–48.
- Ristic, J., & Enns, J.T. (2015). The changing face of attentional development. *Current Directions in Psychological Science*, 24, 24–31.
- Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52.
- Saults, J.S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, 136, 663–684.
- Sek, A., & Moore, B.J. (1995). Frequency discrimination as a function of frequency, measured in several ways. *Journal of the Acoustical Society of America*, 97, 2479–2486.
- Shepard, R.N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305–333.
- Simmering, V.R., & Miller, H.E. (2016). Developmental improvements in the resolution and capacity of visual working memory share a common source. *Attention, Perception, and Psychophysics*, 78, 1538–1555.
- Stevanovski, B., & Jolicoeur, P. (2007). Visual short-term memory: Central capacity limitations in short-term consolidation. *Visual Cognition*, 15, 532–563.
- Tenenbaum, J.B., Kemp, C., Griffiths, T.L., & Goodman, N.D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331, 1279–1285.
- Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1072–1095.
- Vogel, E.K., McCollough, A.W., & Machizawa, M.G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438, 500–503.
- Wallis, G., Stokes, M., Cousijn, H., Woolrich, M., & Nobre, A.C. (2015). Frontoparietal and cingulo-opercular networks play dissociable roles in control of working memory. *Journal of Cognitive Neuroscience*, 27, 2019–2034.

How to cite this article: Cowan N, Li Y, Glass B, Scott Saults J.

Development of the ability to combine visual and acoustic information in working memory. *Dev Sci*. 2018;21:e12635.

<https://doi.org/10.1111/desc.12635>