Working Memory Develops at a Similar Rate Across Diverse Stimuli

Fabien Mathy  
Université Côte d’Azur  

Ori Friedman  
University of Waterloo

Children’s working memory improves with age. We examined whether the rate of improvement varies across different classes of stimuli, or is instead constant across classes of stimuli. We tested between these two possibilities by having participants ($N = 99$) from four age groups (7, 9, 11, adults) complete simple span tasks using items from six stimulus classes. Participants’ span improved with age, and varied across the different stimulus classes. Crucially, age-related improvements were similar across the different stimulus classes. These findings suggest that age-related improvements in working memory result from an increase in capacity, and not from gains in the ability to form chunks or from growing familiarity with certain classes of stimuli. Moreover, the findings build on previous studies on adults showing that working memory performance varies across different stimulus classes by revealing that these differences occur in young children, and remain stable across development.

People can only hold a very limited amount of information for immediate recall. For example, working memory in adults is estimated to have a capacity of just $4 \pm 1$ chunks of information (Cowan, 2001). This capacity might seem unrealistically limited, as we sometimes appear to hold far more information in mind. Effectively, chunks can vary in size, allowing considerable information to be stored using just four chunks (Mathy & Feldman, 2012). For example, the long series of English letters CIAABCFAQPDF can be remembered by forming the chunks CIA, ABC, FAQ, and PDF. The limit of 4 chunks can also seem too generous, as we sometimes have difficulty remembering even just a couple of letters from a foreign alphabet (Alvarez & Cavanagh, 2004). Such difficulties may arise because some stimuli require more than one chunk of information to be correctly encoded with all their features intact (Hardman & Cowan, 2015). Regardless, these examples show that immediate recall varies across stimulus classes (e.g., familiar letters vs foreign ones).

Working memory also varies across the lifespan, and steadily improves through childhood (Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Case, Kurland, & Goldberg, 1982; Dempster, 1981; Pascual-Leone, 1970). However, we know relatively little about whether this improvement depends on the stimuli to be remembered. To explore this issue, we test whether developmental improvement in capacity varies across different classes of stimuli. For example, we examine whether children show greater memory gains for familiar than foreign letters. Addressing this question can provide a window into the nature of age-related improvements in immediate memory.

Variation Across Stimulus Classes

One general possibility is that improvement in immediate recall differs across stimulus classes. If children’s ability to form chunks increases with age, their memory might improve greatly for stimulus classes that are simple, as greater chunking ability should allow older children to store these stimuli using fewer (but larger) chunks. In contrast, memory improvements might be relatively slower for complex stimulus classes, as these stimuli might be difficult to chunk even as children’s ability to form chunks improves.

On this view, we should expect a strong interaction between age and stimulus class. For example, suppose young children can remember 4 objects from a simple stimulus class or 2 from a complex class (i.e., as might occur if 2 simple objects can be grouped into a single chunk; see Brady, Konkle, & Alvarez, 2009). If there are greater developmental gains in the ability to recode simple stimuli than complex ones, then adults might remember 7 objects from the simple class but only 3 from the complex class. That is, the ratio in memory across different classes of items should differ with age (i.e., $4/2 \neq 7/3$).

Differing rates of improvement across stimulus classes might also be expected if increases in children’s immediate memory depend on knowledge of stimuli or familiarity with them, as greater familiarity with stimuli might allow children to store them using fewer chunks. Such effects of knowledge and familiarity have been shown in adults (Jones & Macken,
2015; Reder, Liu, Keinath, & Popov, 2016), and developmental studies suggest that increases in memory span may depend on whether the material allows children to rely on these two factors (e.g., Chi, 1978; Jones & Macken, 2018; Cowan, 2016). For instance, Jones and Macken (2018) show an influence of long-term linguistic influence on short-term memory tasks.

**Uniformity Across Stimulus Classes**

Alternatively, improvements in immediate memory might be uniform across diverse stimulus classes. This prediction follows from accounts holding that developmental improvements in immediate memory chiefly stem from increases in the number of slots available to retain information (Burtis, 1982; Cowan, Ricker, Clark, Hinrichs, & Glass, 2015; Gilchrist, Cowan, & Naveh-Benjamin, 2009). On this view, we should expect only a weak interaction between age and stimulus class.

We can illustrate this by revisiting the example where young children can remember 4 objects from a simple stimulus class or 2 from a complex class. If improvements in memory performance primarily stem from an increase in the number of chunks available, then if adults can remember 4 complex objects, they should be able to remember 8 simple ones. That is, the ratio in memory across different classes of items should remain constant with age (i.e., $4/2 = 8/4 = 2$). This idea applies independently of the number of chunks actually required to encode various objects. Regardless of the actual number, if changes in memory span chiefly reflect an increase in the number of chunks available, then the rate of improvement should be similar across varied classes of stimuli.

Some previous findings are consistent with this prediction. Cowan (2016, Figure 1, using data from Gathercole, Ambridge, Wearing, & Pickering, 2004) shows findings that suggest uniform rates of development between ages 4 and 15 across several kinds of stimuli, including digits, words, and non-words (but not spatial configurations, which develop at a different rate). For example, the spans for nonwords are approximately 1.5 at age 4 vs 2.9 at age 15, and the spans for digits are respectively 3.2 vs 5.8. The developmental ratios (2.9/1.5 and 5.8/3.2) are both around 2. However, these stimulus classes might not differ substantially in terms of their information content. So perhaps greater variations would be observed if a more diverse set of classes was examined.

The prediction that improvements in immediate memory is uniform across diverse stimulus classes also follows from claims that working memory capacity depends on a shared continuous resource (Bays, Catalao, & Husain, 2009). Greater resources let working memory encode material with greater precision, and precision has been shown to improve with age (Simmering & Patterson, 2012). From the perspective of such models, capacity can be thought as one unique large ‘slot’ containing a fixed amount of resource at a given age for encoding stimuli. Imagine this resource amounts to 16 units (e.g., bits of information) letting individuals to encode 4 objects each requiring 4 units or just one object requiring greater precision, that is consuming 16 units on its own. If this resource doubles with age, the same individual could then encode 8 objects each requiring 4 units or 2 objects of 16 units. The ratios would again be constant in this example.

**The Current Experiment**

We tested whether improvements in immediate memory vary or are stable across stimulus classes by capitalizing on Alvarez and Cavanagh’s (2004) finding that immediate memory in adults differs across stimulus classes. For example, they found that adults have greater memory for drawings of familiar objects than for random polygons and cubes with different sides shaded. For this particular set, adults could retain almost three times as many colors, or about twice as many letters as polygons; see Figure 1 for all six stimulus classes used in their study. This set of classes of stimuli therefore seems to provide sufficiently important differences in information content to detect different age-related improvements. For instance, although Alvarez and Cavanagh focused on adults, we used their stimuli with participants in four age groups (7 years, 9 years, 11 years, adults).

We tested participants using simple span tasks (Brainer, 1940; Jacobs, 1887; Unsworth & Engle, 2006). In such tasks, participants are shown a series of stimuli in serial order, and then attempt to recall them in the correct order immediately afterwards; success with longer series indicates higher span. Crucially, across testing trials, we varied the class of stimuli in the to-be-remembered series. This allowed us to test whether the span at each group varied across stimulus classes, and whether improvements in immediate memory vary across classes or are instead constant.

We decided against using a change detection task (i.e., as employed by Alvarez and Cavanagh) because previous studies have provided mixed results with children. Whereas some authors (Cowan et al., 2015) found similar rate of developmental change across familiarity levels, others have not (Sørensen & Kyllingsbæk, 2012). Also, because we wanted to give children ample opportunity to encode the materials, we chose not to use a concurrent task, like in the complex span task. In any case, complex span tasks and simple span tasks are thought to measure the same basic processes (Colom, Shih, Flores-Mendoza, & Quiroga, 2006; Martínez et al., 2011), particularly when list lengths increase (Unsworth & Engle, 2007).

Some properties of simple span tasks make them especially attractive for testing children: 1) These tasks have provided reliable estimates of immediate memory since the 1920s, and show resistance to the Flynn effect (Gignac,
WORKING MEMORY ACROSS STIMULI

Figure 1. Stimuli used by Alvarez and Cavanagh (2004) and in the present experiment. From top to bottom: colors, alphabet, objects, kanjis, polygons, cubes.

2015). As such, the cognitive processes they assess appear to be central to cognition. 2) The procedure is relatively slow, and this helps avoid issues stemming from momentary lapses of attention. This slowness also increases opportunities for children to encode stimuli, which is important given that some stimuli may become more meaningful with increased age and experience. 3) Serial order may prevent older children from using strategies that are unlikely to be available to younger children. For example, because the task requires children to reconstruct order, it prevents the strategy available in free recall tasks of recalling the last items first, then the first ones, then guessing the middle items (see Chi, 1978 who made this observation in adults).

Method

Participants. We established a minimum target number of participants per age group as follows: We expected that consecutive age group would differ by a span of one object on average, with a standard deviation of one. Power analysis targeting a power of .90 to compare 4 means with a difference of 1 on average (i.e., 3 successive pairwise comparisons, 1-sided, pooled sd = 1, balanced design) produced a minimal sample size of 24 per group. However, when our recruitment efforts yielded some additional children in some age groups, we tested all the children for whom permission was obtained up to 25. We tested 99 participants in four age groups. These were 25 7-year-olds (M = 6.8 years, SD = .37), 22 9-year-olds (M = 8.8 years, SD = .31), 27 11-year-olds (M = 10.9 years, SD = .34), and 25 young adults (M = 22, SD = 1.76). The children were from middle-class families, and were recruited from, and tested at a public school. The young adults were enrolled at the University XXX. The children were tested in a quiet area outside of the class environment.

Procedure. Participants completed a series of trials, in which they saw a series of items and then attempted to remember the items in their correct order. In each trial, stimuli from Alvarez and Cavanagh (2004; see Figure 1) were presented one-at-a-time at the center of a computer display for 1000 ms each. After all items for the trial had been presented, a fixation cross appeared for 1000 ms. Then, all items from the trial simultaneously appeared on the display in a random arrangement, and participants were required to indicate the items in the order in which they originally appeared. Participants indicated items using a computer mouse.

Each participant completed 36 trials, yielded by crossing the six classes of stimuli with trial lengths ranging from 1 to

---

1These estimates were derived by averaging a maximal expected difference between 7-year-olds and adults of 4 (based on Pascale Leone, 1970), and a minimal expected difference of 2 (based on Cowan, 2001). This yields an average difference of 3, and so we expected a difference of 1 between consecutive age groups.
6 items. Trials were blocked by stimulus class, and were ordered within blocks from shortest (1 item) to longest (6 items). The order of the stimulus class blocks was random.

**Scoring.** In the main analyses, participants were scored using the all-or-nothing method (Conway et al., 2005), in which participants were assigned 1 point for each perfectly correct serial report. Participants had separate scores ranging from 1 to 6 for each of the six stimulus classes. However, as reported below, in further analyses we also used partial-credit scoring.

**Results**

The data are available at [https://osf.io/p5uqc/?view_only=c51b5cb1d747210c2914e62a8463a853](https://osf.io/p5uqc/?view_only=c51b5cb1d747210c2914e62a8463a853). Participants' average spans for each stimulus class are shown in Fig. 2. Span scores were first analyzed using a repeated-measures ANOVA with the between-subjects factor Age-group (7, 9, 11, and Adults) and the repeated-measures factor Stimulus class (alphabet, colors, objects, kanjis, cubes, polygons). This analysis revealed main effect of Age-group, in which scores increased with age, \( F(3, 95) = 60.10, p < .001, \eta_p^2 = .66 \). There was also a main effect of Stimulus class, \( F(5, 475) = 54.95, p < .001, \eta_p^2 = .37 \), but no interaction between Age-group and Stimulus class (\( p = .10 \)).

To follow-up on the main effect of Stimulus class, we examined how spans varied across the different stimulus classes. From greatest to smallest, the mean spans were: alphabet (M = 4.58), colors (M = 4.03), objects (M = 3.90), Kanjis (M = 3.22), cubes (M = 2.73), and polygons (M = 2.64). Post-hoc pairwise comparisons using Bonferroni correction revealed that, aside from two exceptions, all spans significantly differed from one another, all \( p < .05 \) with a \( d \) at least superior to .43. The two exceptions were no significant differences between polygons and cubes (\( p > .999 \)), and no significant difference between objects and colors (\( p > .999 \)). This variation between spans is similar to that observed by Alvarez and Cavanagh (2004), who obtained respective spans of 3.7, 4.4, 2.6, 2.8, 1.6, and 2.0, as the correlation between our mean spans and theirs is .82 (\( N = 6 \)). However, there were some notable differences in the relative ordering of spans between the experiments, and we observed longer spans, particularly in adults. In the Discussion, we further consider similarities and differences between our findings and those of Alvarez and Cavanagh (2004).

To follow-up on the finding of no interaction between Age-group and Stimulus class, we conducted two further tests to confirm that differences between spans for the different stimulus classes did not change with age. First, we repeated the main analysis using a Bayesian repeated-measures ANOVA. To conduct this analysis, we used JASP (retrieved from http://jasp-stats.org/) with default parameters. This analysis found that stimulus type and age group combined showed clear evidence against the null model (\( BF_{10} = 5.06^{10} \)). The Bayes factor against a model including the interaction term was 13 (\( 5.06^{10} / 3.77^{19} = 13 \)). This provides further evidence against an interaction between age and stimulus class. The inclusion Bayes factor (i.e., the change from prior to posterior inclusion odds) was equal to 0.5 for the interaction, against infinite values for stimulus type and age group.

Second, to provide an even stronger test of additivity of age and stimulus class, we examined the ratios between the different spans at each age. For these analyses, we averaged the spans of the polygon and cube classes, and the object and color classes, as our post-hoc tests revealed no difference between these classes. After collapsing, we had 4 estimates of the span of each participant (i.e., estimates for alphabet, objects-colors, Kanjis, and polygons-cubes). We then computed 6 span ratios (one for each possible pairing of spans): Kanjis:polygons-cubes, objects-colors:polygons-cubes, alphabet:polygons-cubes, objects-colors:Kanjis, alphabet:Kanjis, and alphabet:objects-colors. For each of the 6 ratios, we used JASP with default parameters to run a simple Bayesian ANOVA with Age-group (7, 9, 11, and Adults) as a between-subjects factor. For each of these 6 Bayesian ANOVAs, the null hypothesis was that the ratios were equal across age groups, and the alternative hypothesis was that the ratios varied across age groups. In all 6 analyses, the result was in favor of the null. The 6 Bayes factors (in favor of the null) were 3 (Kanjis/polygons-cubes), 13 (objects-colors/polygons-cubes), 32 (alphabet/polygons-cubes), 2 (objects-colors/Kanjis), 2 (alphabet/Kanjis), and 3 (alphabet/objects-colors). These analyses show that the ratios between different spans are constant across age, and that age-related improvements in memory are constant across different classes of stimuli.

**Alternative analytic approaches**

We ran further analyses in which we normalized the spans by computing a \( z \) score around the overall mean and standard deviation (see Cowan et al., 2015). Using the \( z \) scores, the growth curves coincided and the repeated-measures ANOVA did not reach significance for the stimulus-class factor (\( F = .006 \) and \( p = 1 \)); the interaction was still nonsignificant and age was still significant.\(^2\)

\(^2\)With list lengths of 1, the test was not meaningful. This list length served as a warm-up to let participants be more confident with the task.

\(^3\)Although the relationship between age and stimulus class could appear strictly additive using a log scale, we did not use a log scale to transform the data. The main reason is that a log scale might not apply to all cases similarly. Also, this method could eventually flatten the numbers artificially. We thought it was best to compute the ratios for each participant separately because the ratios directly describe the potential gain from one stimulus class to another, or to use the \( z \) scores as a more general method.
Although the main analyses used all-or-none scoring, we also examined the data using partial-credit scoring, in which we summed the number of items in each trial recalled at their correct position. These partial-credit scores strongly correlated with the all-or-nothing scores aggregated across participants and types of stimuli (Pearson correlation between the two methods: $r = .92, N = 594, p < .001$). As a result, the findings were similar to the main analysis based on the all-or-nothing scores when we reran the repeated-measures ANOVA on the partial-credit scores.

**Discussion**

We examined children’s and adults’ immediate recall for six stimulus classes, and observed three main findings. First, immediate recall improved with age. Our youngest participants, who were aged around 7 performed worst; children at ages 9 and 11 performed progressively better; and adults performed best of all. Second, participants’ immediate memory performance varied across the different stimulus classes. Third, although memory performance varied across the stimulus classes, the age-related improvements in span were similar across the different stimulus classes, meaning that all stimulus classes yielded a similar increase in capacity with age. These findings are informative about the development of working memory capacity and the mechanisms underlying age-related improvements in memory performance.

Our findings are consistent with accounts claiming that there is fundamental development growth in working memory capacity (e.g., Burtis, 1982; Cowan et al., 2015; Gilchrist et al., 2009). This growth could result from an increase in the number of discrete slots (Rouder et al., 2008) or from an increase in the magnitude of a shared continuous resource that can be allocated to objects (Bays et al., 2009). Both explanations (Donkin, Tran, & Nosofsky, 2014) seem equally compatible with our results. In accounts positing discrete slots, fewer objects can be recalled if each object is complex and requires several slots to be encoded properly. In accounts positing a continuous resource, fewer objects can be recalled when the objects are complex and require greater allocation of the resource.

In both accounts, age-related changes should lead to similar memory increases for simple and complex stimuli alike.
Consistent with this, our participants showed a gain in capacity of 50% between age 7 and adulthood, irrespective of stimulus class. However, these results differ from those of Cowan (2016) who rescored data from Gathercole et al. (2004), and found steeper development for stimuli, like visual patterns, that can be potentially recoded to form spatial configurations.

In contrast, the present findings do not fit well with the idea that developmental increases in working memory capacity stem from improvements in the ability to form chunks or from developmental increases in familiarity with stimuli from certain classes (for a recent review of this question, see Cowan, 2016). If either of these factors was responsible for age-related improvements in immediate memory, then improvements should have varied across the different classes of stimuli. It is important to note, though, that we are not suggesting that these two factors remain static across development. The ability to form chunks may improve with age, and it is likely that there are age-related gains in familiarity and knowledge for certain stimulus classes. Based on our findings, though, these developmental changes do not contribute to age-related improvements in working memory (at least as assessed by simple span tasks, which do not provide much time to encode items).

Our findings also build on Alvarez and Cavanagh (2004)’s findings that working memory capacity varies across different stimulus classes. Our findings show that these differences occur in young children, and not just in adults. Nonetheless, there were some differences between our experiment and theirs in the relative ordering of memory spans for the stimulus classes. These differences between our experiment and theirs are unlikely to stem from differences in the ages of the participants between the experiments (i.e., we tested children and adults, whereas they only tested adults), as we found that memory performance was not subject to an age by stimulus class interaction. It is more likely that some performance differences resulted because we used a simple span task, whereas they used a change detection task. Even so, these differences are relatively minor. Overall, the experiments agree on which stimulus classes support greater and lesser immediate recall performance.

Although our findings are informative about the development of working memory across much of childhood, our youngest participants were aged around 7 years. This means that we cannot be sure that our conclusions extend to younger children, such as preschoolers. Hence, it remains possible that preschoolers’ improvements in immediate memory do stem from improvements in chunking ability (Kibbe & Feigenson, 2014). This said, younger children might already show similar chunking abilities to our participants as the ability to form more efficient representations by chunking is already present in infants (Kibbe & Feigenson, 2016).

Finally, it would be useful for future studies to determine whether there is a fundamental developmental difference between simple span tasks, complex span tasks, and tasks using the change detection paradigm, for the stimuli employed in the current study. We suspect the mostly likely difference that might emerge with complex or change detection tasks would be a narrower range of spans across stimulus classes at each age, because these tasks typically yield lower span estimates than simple span tasks. (Crucially, this difference is non-developmental and would not affect our conclusion that development is uniform across stimulus classes.) If complex or change detection tasks did reveal a narrower range of spans across stimulus classes, this would suggest that the substantial variation across classes that we observed resulted from chunking. Alternatively, if other tasks revealed a similarly broad range of spans at each age, this would disconfirm that chunking is responsible for the variations observed across stimulus classes in simple span tasks and this could mean that familiarity (or other factor) with stimuli is a more determinant factor across tasks.
References


Gignac, G. E. (2015). The magical numbers 7 and 4 are resistant to the Flynn effect: No evidence for increases in forward or backward recall across 85 years of data. *Intelligence, 48*, 85-95.


