Working memory develops at a similar rate across diverse stimuli

Fabien Mathy [a,*], Ori Friedman [b]

[a] Centre National de la Recherche Scientifique (CNRS), Laboratory UMR 7320, Université Côte d'Azur, 06100 Nice, France
[b] Department of Psychology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

ARTICLE INFO

ABSTRACT

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 years, 9 years, 11 years, and 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 mostly 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.

Introduction

In adults, immediate recall varies substantially across stimulus classes. For example, adults can typically remember more letters from their own alphabet than letters from a foreign one, and they can also remember more distinct colors than multisided polygons (Alvarez & Cavanagh, 2004).

---

* Corresponding author.
E-mail address: fabien.mathy@univ-cotedazur.fr (F. Mathy).

Although there are no systematic effects found in the literature, such variations in adults' memory could result from differences in familiarity between items (Chen, Eng, & Jiang, 2006; Hulme, Maughan, & Brown, 1991; Pashler, 1988; Reder, Liu, Keinath, & Popov, 2016), from differences in the complexity or regularity of items (Awh, Barton, & Vogel, 2007; Brady & Tenenbaum, 2013), and possibly from variations in adults' ability to chunk or compress items (Brady, Konkle, & Alvarez, 2009; Norris, Kalm, & Hall, 2019). Prior knowledge and familiarity could positively influence the fidelity with which the items are stored (Hemmer & Steyvers, 2009), and these factors could eventually influence working memory capacity. Research has long focused on the number of items that best defines capacity limits (Cowan, 2001; Miller, 1956), but during recent years there has been increased focus on the nature of the stored items (for a review, see Brady, Konkle, & Alvarez, 2011).

Immediate recall also varies across the lifespan and steadily improves throughout 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 nature of the stimuli to be remembered. To explore this issue, we tested whether developmental improvement in capacity varies across different classes of stimuli. For example, we examined whether children show greater memory gains for familiar letters than for foreign letters. Addressing this question can provide a window into the nature of age-related improvements in immediate memory. If capacity varies with the nature of the stored items early on during childhood, this could question whether adults have the privilege of knowledge to inform immediate recall.

Variation in rates of improvement 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 because 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 because 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 that young children can remember 4 objects from a simple stimulus class or 2 objects from a complex class (i.e., as might occur if 2 simple objects can be grouped into a single chunk; see Brady et al., 2009). If there are greater developmental gains in the ability to recode simple stimuli than complex stimuli, then adults might remember 7 objects from the simple class but only 3 objects from the complex class. That is, the ratio in memory across different classes of items should differ with age (i.e., $4/2 \neq 7/3$), corresponding to two different rates of development (i.e., 50% for the complex class's increase from 2 to 3 objects and 75% for the simple class's increase from 4 to 7 objects).

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 because 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 et al., 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; Cowan, 2016; Jones & Macken, 2018). For instance, Jones and Macken (2018) showed an influence of long-term linguistic influence on short-term memory tasks.

Uniformity in rate of improvement across stimulus classes

Alternatively, immediate memory might improve at a uniform rate 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). We can illustrate this by revisiting the example where young children can remember 4 objects from a simple stimulus class or 2 objects from a complex class. If improvements in working memory performance primarily stem from an increase in the number of chunks (or slots) available, then if adults can

remember 4 complex objects, they should be able to remember 8 simple objects. That is, the ratio in memory across different classes of items should remain constant with age (e.g., 4/2 = 8/4 = 2, also called a multiplicative effect). This idea applies independent 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 (e.g., a constant proportional growth of 100% for 4/2 and 8/4). On this view, we can expect a "weak" interaction between stimulus class and age.[1]

Some previous findings are consistent with this prediction. Cowan (2016, Fig. 1, using data from Gathercole, Ambridge, Wearing, & Pickering, 2004) showed findings that suggest uniform rates of development between 4 and 15 years of age across several kinds of stimuli, including digits, words, and nonwords (but not spatial configurations, which develop at a different rate). For example, the spans for nonwords are approximately 1.5 at 4 years of age versus 2.9 at 15 years of age, and the spans for digits are 3.2 versus 5.8, respectively. The developmental ratios (2.9/1.5 and 5.8/3.2) both are 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 are 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 of as one unique large "slot" containing a fixed amount of resource at a given age for encoding stimuli. Imagine that this resource amounts to 16 units (e.g., bits of information), letting individuals encode 4 objects each requiring 4 units or just 1 object requiring greater precision, that is, consuming 16 units on its own. If this resource doubles with age, then the same individual could 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 (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 nearly three times as many colors or about twice as many letters as polygons (see Fig. 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. Whereas Alvarez and Cavanagh focused on adults, we used their stimuli with participants in four age groups (7 years, 9 years, 11 years, and adults).

We tested participants using simple span tasks (Brener, 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 afterward; success with longer series indicates higher span. Crucially, across testing trials, we varied the class of stimuli between the to-be-remembered series. This allowed us to test between the two accounts. Specifically, it allowed us to test whether the ratios in memory across stimulus classes varied with development or instead remained uniform.

We decided against using a change detection task (i.e., as employed by Alvarez & Cavanagh, 2004) because previous studies have provided mixed results with children. Whereas some authors have found similar rates of developmental change across familiarity levels (Cowan et al., 2015), others have not (Sørensen & Kyllingsbæk, 2012). In addition, we wanted to give children ample opportunity to encode the materials.

---

[1] A third possibility is that immediate memory improves to a similar degree across diverse stimulus classes. For example, if memory for complex stimuli improved from two to four items, then memory for simple ones would improve from four to six items (i.e., again of two items for each class). This additive effect predicts no interaction between age and stimulus class; it predicts that improvements in immediate memory should be parallel across stimulus classes.
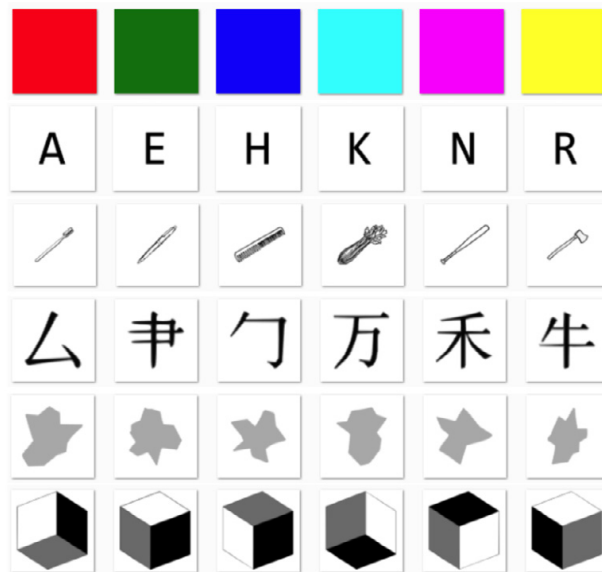
Fig. 1. Stimuli used by Alvarez and Cavanagh (2004) and in the current experiment. From top to bottom: Colors, alphabet, objects, kanjis, polygons, and cubes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Some properties of simple span tasks make them especially attractive for testing children. First, these tasks have provided reliable estimates of immediate memory since the 1920s and show resistance to the Flynn effect (Gignac, 2015). As such, the cognitive processes they assess appear to be central to cognition. Second, the procedure is relatively slow, and this helps to 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. Third, 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 recalling the first items, and 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 groups would differ by a span of 1 object on average with a standard deviation of 1.[2] Power analysis targeting a power of .90 to compare 4 means with a difference of 1 on average (i.e., three successive pairwise comparisons, one-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 = 82 months, SD = 4.4, range = 76–91), 22 9-year-olds (M = 106 months, SD = 3.8, range = 101–115), 27 11-year-olds (M = 131 months, SD = 4.1, range = 125–142), and 25 young adults (M = 265 months, SD = 21.2, range = 231–310). The children were from middle-class families and were recruited from and tested at a public school. The children were tested in a quiet area outside of the classroom environment. The young adults were enrolled at the

---

[2] These estimates were derived by averaging a maximal expected difference between 7-year-olds and adults of 4 (based on Pascual-Leone, 1970) and a minimal expected difference of 2 (based on Cowan, 2001). This yields an average difference of 3, so we expected a difference of 1 between consecutive age groups.

University Côte d'Azur and were tested in a lab in a room dedicated to running experiments with single individuals.

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 Fig. 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 6 items.[3] 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. We limited the number of trials to 36 to ensure that the task, including instructions, would take 10–15 min (i.e., increasing the likelihood that children would remain attentive and motivated).

Scoring

In the main analyses, participants were scored using partial credit unit scoring, in which we computed the proportion of items recalled at their correct position per sequence and then summed scores across sequences. For instance, if a participant responded perfectly for the 1- to 5-item sequences and then made two errors in the 6-item sequence, the participant's span would be 1 + 1 + 1 + 1 + 1 + .66 = 5. 66 (see Conway et al., 2005, for discussion and findings favoring this approach of scoring over alternatives).

Results

The data are available at https://osf.io/p5ujc/?view_only=c51529eb1477421ca914c62d844b3853. To conduct the analyses, we used JASP (retrieved from http://jasp-stats.org/) with default parameters.

Participants' average spans for each stimulus class are shown in Fig. 2. Span scores were first analyzed using a standard repeated-measures analysis of variance (ANOVA) with the between-participants factor age group (7 years, 9 years, 11 years, or adults) and the repeated-measures factor stimulus class (alphabet, colors, objects, kanjis, cubes, or polygons). This analysis revealed a main effect of age group, $F(3, 95) = 54.00$, $p < .001$, $\eta_p^2 = .63$. Post hoc tests showed that the only nonsignificant difference between age groups occurred between 9 and 11 years. There was also a main effect of stimulus class, $F(5, 475) = 71.54$, $p < .001$, $\eta_p^2 = .43$, and an interaction between age group and stimulus class, $F(15, 475) = 2.22$, $p = .005$, $\eta_p^2 = .07$. Although the interaction had a small effect size, a Bayesian repeated-measures ANOVA confirmed that the full model (i.e., both factors including the interaction term) received evidence against the null model ($BF_{10} = 1.7e+69$) and that including the interaction term increased the model probability $(BF_M = 12.7)$.

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 ($M = 4.83$), colors ($M = 4.43$), objects ($M = 4.40$), kanjis ($M = 3.77$), cubes ($M = 3.48$), and polygons ($M = 3.32$). Post hoc pairwise comparisons using Bonferroni correction revealed that, aside from two exceptions, all spans significantly differed from one another (all $ps < .05$) with a d at least superior to .30. The two exceptions were no significant differences between polygons and cubes ($p = .66$) and no significant difference between objects and colors ($p > .999$). There were some notable differences in the relative ordering of our spans and those observed by Alvarez and Cavanagh (2004), who obtained spans of

---

[3] 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.
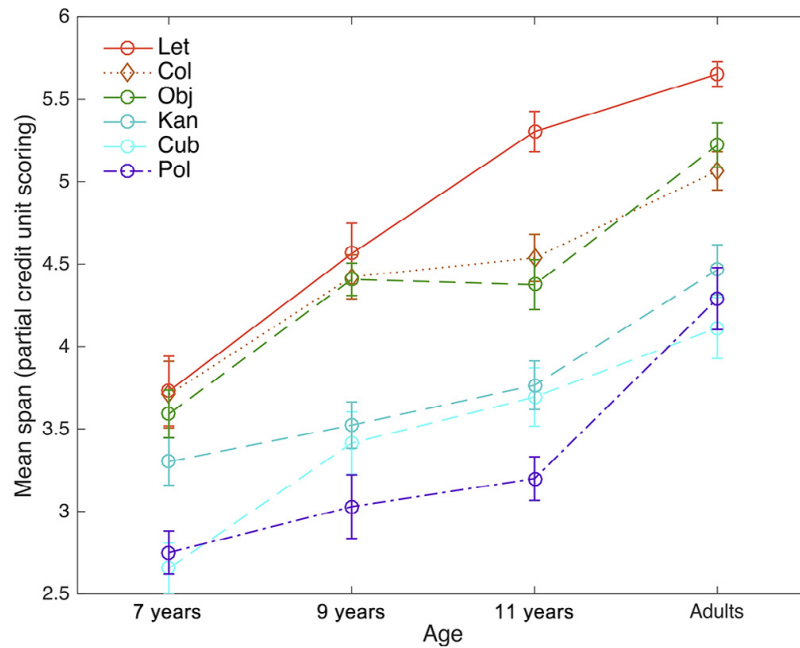
Fig. 2. Mean spans across stimulus classes and age groups. Let, alphabet; Col, colors; Obj, objects; Kan, kanjis; Cub, cubes; Pol, polygons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.7, 4.4, 2.6, 2.8, 1.6, and 2.0, respectively. In the Discussion, we further consider similarities and differences between our findings and those of Alvarez and Cavanagh.

We next followed up on the interaction between age group and stimulus class. As noted in the Introduction, such an interaction could result from different stimulus classes improving at different rates, or it could result from different classes all improving at a uniform rate. Furthermore, the interaction could also result from inconsistent fluctuations in memory performance. To test among these three possibilities, we examined whether the ratio in memory across different classes of items varied with age.

For this ratio analysis, we averaged the spans of the polygon and cube classes, and the object and color classes, because our post hoc tests revealed no difference between these classes. This averaging was intended to reduce the number of paired comparisons. After collapsing, we had four estimates of the span of each participant (i.e., estimates for alphabet, colors–objects, kanjis, and cubes–polygons). We then computed six span ratios (one for each possible pairing of spans): alphabet/colors–objects, alphabet/kanjis, alphabet/cubes–polygons, colors–objects/kanjis, colors–objects/cubes–polygons, and kanjis/cubes–polygons. For each of the six ratios, we used JASP with default parameters to run a simple Bayesian linear regression with age (in months) as the covariate; we treated age continuously to increase the power for this analysis. For each of these six Bayesian tests, the null hypothesis was that the ratios were equal across age, and the alternative hypothesis was that the ratios would progress with age. The results overall favored the null rather than the alternative. The values of the six Bayes factors in favor of the null (BF_{01}) were 1.3 (kanjis/cubes–polygons), 0.9 (colors–objects/cubes–polygons), 3.4 (alphabet/cubes–polygons), 4.4 (colors–objects/kanjis), 4.2 (alphabet/kanjis), and 3.5 (alphabet/colors–objects). Thus, four Bayes factors showed moderate evidence in favor of the null (i.e., values > 3) and none showed evidence in favor of the alternative.

Analyses without the adult group

A potential concern with the preceding analysis is that adults could have shown ceiling effects because the list lengths were limited to 6 items. To address this concern, we reran the analyses without the adults.

Without adults, we observed similar results; the standard ANOVA again revealed significant main effects and an interaction between age group and stimulus class, $F(5, 360) = 4.44$, $p < .001$, $\eta_p^2 = .06$. The Bayesian repeated-measures ANOVA again confirmed that the full model (i.e., both factors including the interaction term) received evidence against the null model ($BF_{10} = 2.8e+43$) and that including the interaction term increased the model probability ($BF_M = 19.5$).

The six Bayesian linear regression analyses of the increase of the span ratios as a function of age (in months) entered as a covariate, however, gave the following Bayes factors in favor of the null (BF_{01}): 2.2 (kanjis/cubes–polygons), 3.2 (colors–objects/cubes–polygons), 1.5 (alphabet/cubes–polygons), 3.1 (colors–objects/kanjis), 0.2 (alphabet/kanjis), and 0.4 (alphabet/colors–objects). Thus, only one Bayes factor showed evidence in favor of the alternative hypothesis, namely, for the alphabet/kanjis ratio ($BF_{10} = 4.6$).

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 of each stimulus class across all age groups (see Cowan et al., 2015). This method is an intuitive approach because the growth curves perfectly coincide when the ratios are constant.[4] Using the z scores, the six growth curves coincided and the Bayesian repeated-measures ANOVA indicated that the interaction term did not show sufficient evidence to be integrated in the model ($BF_M = 0.002$), and there was very strong evidence that there was no effect of the stimuli ($BF_{01} = 780$). A similar result was obtained by removing the adult participants from the analysis ($BF_M = 0.01$ and $BF_{01} = 260$).

Overall, these analyses show that age-related improvements in memory are mostly similar across different classes of stimuli.

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 years, performed worst, children at 9 and 11 years of age performed better, and adults performed best of all. Second, participants' immediate memory performance varied across the different stimulus classes. Third, the rate of improvement in immediate memory was mostly uniform across stimulus classes, as revealed by our analysis of the ratios between stimulus classes. These findings are informative about the development of working memory capacity and the mechanisms underlying age-related improvements in memory performance.

Our findings are generally 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 to be 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 a similar rate of memory increase for simple and complex stimuli alike. Consistent with this, our data could be roughly described as participants showing a proportional gain in capacity of about 50% between 7 years of age and adulthood irrespective of stimulus class. This said, it is important to acknowledge that our results differ from those of

---

[4] Alternatively, we could have used a logarithm to transform the data. Effectively, a multiplicative relationship between two factors becomes additive using a log scale, so it can detect proportional growths (Kerkhoff & Enquist, 2009). However, we thought that it was best to compute the ratios for each participant and each pair of stimulus classes to describe the participants' potential gains from one stimulus class to another separately, which more directly indicate the size of the chunks than a logarithm. Nevertheless, we used the z score because it also provides an intuitive general result.

Cowan (2016), who rescored data from Gathercole et al. (2004) and found steeper development for stimuli, such as visual patterns, that can be potentially recoded to form spatial configurations.

In contrast, the current 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, the rate of improvement should have varied substantially across the different classes of stimuli. It is important to note, however, 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, however, 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) 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 may have resulted because we used a simple span task, whereas Alvarez and Cavanagh used a change detection task. For instance, the simple span task could have increased verbal encoding. In the current study, all age groups were relatively better at remembering stimulus types that were privy to verbal encoding (i.e., alphabet letters, colors, and objects). This advantage could be positioned in terms of Baddeley (1986) oft-cited distinction between the visuospatial sketchpad and the phonological loop. Even so, the differences between our experiment and that of Alvarez and Cavanagh (2004) 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 because the ability to form more efficient representations by chunking is already present in infants (Kibbe & Feigenson, 2016). It is also possible that findings would differ if children were given more time to encode the items or if the experiment included more trials. Both of these manipulations might facilitate chunking of stimuli, and this could differentially affect memory for some stimulus classes (Bower & Winzenz, 1969). Similarly, given concerns we discussed earlier regarding potential ceiling effects, findings might likewise differ if participants were tested with successively longer lists until they each fell below some criteria of success for all stimulus classes.

To draw more definitive conclusions regarding the nature of the relation between age and stimulus type, it would be helpful to use multitask comparisons. For instance, because it has been shown that short-term memory estimates are more domain specific than working memory estimates (Kane et al., 2004), one could expect larger differences across stimulus classes for simple span tasks. Hence, it would be useful for future studies to determine whether there is a fundamental developmental difference among simple span tasks, complex span tasks, and tasks using the change detection paradigm for the stimuli employed in the current study. We suspect that the most 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 nondevelopmental 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 could mean that familiarity (or other factor) with stimuli is a more determinant factor across tasks. However, because the different tasks also share a large degree of variability (in particular simple and complex span tasks; see Colom, Shih, Flores-Mendoza, & Quiroga, 2006; Martínez et al., 2011; Unsworth & Engle, 2007), comparing the tasks would

most likely require both a larger range of list lengths and a larger sample to detect a potential three-way interaction.

Acknowledgments

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106–111.

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

Baddeley, A. (1986). *Working memory*. New York: Oxford 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.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10), 7.

Bower, G. H., & Winzenz, D. (1969). Group structure, coding, and memory for digit series. *Journal of Experimental Psychology, 80*, 1–17.

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

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

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.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26*, 467–482.

Burtis, P. J. (1982). Capacity increase and chunking in the development of short-term memory. *Journal of Experimental Child Psychology, 34*, 387–413.

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, D., Eng, H. Y., & Jiang, Y. (2006). Visual working memory for trained and novel polygons. *Visual Cognition, 14*, 37–54.

Chi, M. T. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops* (pp. 73–96). Hillsdale, NJ: Lawrence Erlbaum.

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory, 14*, 804–813.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–114. discussion 114–185.

Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth?. *Perspectives on Psychological Science, 11*, 239–264.

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.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63–100.

Donkin, C., Tran, S. C., & Nosofsky, R. (2014). Landscaping analyses of the ROC predictions of discrete-slots and signal-detection models of visual working memory. *Attention, Perception, & Psychophysics, 76*, 2103–2116.

Gathercole, S., Ambridge, B., Wearing, H., & Pickering, S. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177–190.

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.

Gilchrist, A. L., Cowan, N., & Naveh-Benjamin, M. (2009). Investigating the childhood development of working memory using sentences: New evidence for the growth of chunk capacity. *Journal of Experimental Child Psychology, 104*, 252–265.

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science, 1*(1), 189–202.

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685–701.

Jacobs, J. (1887). Experiments on "prehension". *Mind, 12*(45), 75–79.

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1–13.

Jones, G., & Macken, B. (2018). Long-term associative learning predicts verbal short-term memory performance. *Memory & Cognition, 46*, 216–229.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.

Kerkhoff, A. J., & Enquist, B. (2009). Multiplicative by nature: Why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology, 257*, 519–521.

Kibbe, M. M., & Feigenson, L. (2014). Developmental origins of recoding and decoding in memory. *Cognitive Psychology, 75*, 55–79.

Kibbe, M. M., & Feigenson, L. (2016). Infants use temporal regularities to chunk objects in memory. *Cognition, 146*, 251–263.

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, M. Á., & Colom, R. (2011). Can fluid intelligence be reduced to "simple" short-term storage?. *Intelligence, 39*, 473–480.

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.

Norris, D. G., Kalm, K., & Hall, J. (2019). Chunking and redintegration in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0000762.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32*, 301–345.

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

Reder, L., Liu, X., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review, 23*, 271–277.

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

Simmering, V. R., & Patterson, R. (2012). Models provide specificity: Testing a proposed mechanism of visual working memory capacity development. *Cognitive Development, 27*, 419–439.

Sørensen, T. A., & Kyllingsbæk, S. (2012). Short-term storage capacity for visual objects depends on expertise. *Acta Psychologica, 140*, 158–163.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54*, 68–80.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*, 1038–1066.