Contents lists available at ScienceDirect

Cognitive Development

Developmental and individual differences in the precision of visuospatial memory

Ioannis Sarigiannidis [a], Gemma Crickmore [b], Duncan E. Astle [b,*]

[a] Department of Education, University of Cambridge, United Kingdom
[b] Medical Research Council, Cognition and Brain Sciences Unit, Cambridge, United Kingdom

A B S T R A C T

Our ability to retain visuospatial information over brief periods of time is severely limited and develops gradually. In childhood, visuospatial short-term and working memory are typically indexed using span-based measures. However, whilst these standardized measures have been successful in characterizing developmental and individual differences, each individual trial only provides a binary measure of a child's performance—they are either correct or incorrect. Here we used a novel continuous report paradigm, in combination with probabilistic modeling, to explore developmental and individual differences in how likely children were to recall memoranda, and how precisely they could report them. Taking this approach revealed a number of novel findings: (i) a concurrent processing demand negatively impacted upon both of these parameters, increasing the guessing rate and making children less precise; (ii) older children (aged 10–12, $N = 20$) were significantly less likely to guess, but when they did remember the target were no more precise in reporting it than younger children (aged 7–9, $N = 20$); (iii) children's performance on standardized short-term and working memory tasks was significantly associated with both the guessing likelihood, and the precision of target responding, on the continuous report task. In short, we show that continuous report paradigms can offer interesting insight into processes that underlie developmental and individual differences in visuospatial memory in childhood.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Our ability to manage and adapt to new and complex situations relies on our capacity to hold in mind small amounts of relevant information for brief periods. Depending upon the context, it may be necessary to hold in mind novel visuospatial or verbal information, for use in some ongoing task. In typically and atypically developing populations this short-term storage capacity is usually measured using span-based measures, with children attempting to maintain increasingly long sequences of items. Sometimes this maintenance can be alongside a concurrent processing (termed working memory, WM) or in isolation (short-term memory, STM). Performance on each trial is coded as correct or incorrect, and the span length is gradually increased until the child's average performance falls below a certain criterion. Performance up to the discontinuation point is summed to give a measure of the child's capacity, which can be expressed either as raw or age-standardized scores. Performance on tasks that tap STM and WM abilities varies markedly across children, with there being both substantial developmental and individual differences in capacity in childhood.

* Corresponding author at: MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom.
E-mail address: duncan.astle@mrc-cbu.cam.ac.uk (D.E. Astle).

Fig. 1. A depiction of the mixture of possible responses that could be observed with a continuous report task (e.g., remembered and reporting line orientations). The blue distribution indicates a proportion of responses that across trials are clustered around the correct target orientation. The red distribution shows the proportion of responses that are clustered around the correct non-target orientation. The green distribution corresponds to a uniform distribution, which represents the proportion of trials upon which subjects produce a random guess.

There are many underlying mechanisms that might drive both developmental improvements in visuospatial STM/WM performance and the large degree of variability across children of the same age. One possibility is that poorer scores, either due to immaturity or individual differences, stem from low capacity per se. That is, what varies across children is the quantity of items that can be maintained simultaneously (Pascual-Leone, 1970). Alternatively, these differences could be underpinned by variability in the quality with which children can maintain items. Some children may be able to maintain items more precisely in memory (Burnett Heyes, Zokaei, van der Staaij, Bays, & Husain, 2012), perhaps reducing the extent to which items interfere with one another or mitigating the impact of decay. A final mechanism that might drive these differences is the extent to which children confuse the order/location of the items that they are storing in memory, leading to a misbinding error—that is, they report an incorrect, but successfully maintained item (Cowan, Naveh-Benjamin, Kilb, & Saults, 2006). Of course, these accounts are not mutually exclusive, and differences in visuospatial memory ability across children could stem from a combination of these mechanisms. A similar set of possible mechanisms could account for the impact of concurrent processing on maintenance—it could reduce the quality of item representation, increase the likelihood of an item being lost from memory, or increase the likelihood of an incorrect item being reported. However, the methods typically used to establish capacity differences across children do not enable us to tease apart these potentially separate underlying component processes. This is because each trial only provides a single binary score that essentially combines all of these potentially separate parameters.

In the adult literature, spurred by an on-going debate as to the nature of resource allocation within visual STM, a number of researchers have started using continuous report as a means for exploring STM processes (Bays & Husain, 2008; Wilken & Ma, 2004; Zhang & Luck, 2008). Participants are presented with memoranda that can be varied in a continuous way, such as hue (Zhang & Luck, 2008) or line orientation (Bays & Husain, 2008), and by providing participants with a means of freely recalling a cued item (for instance using a color wheel or dial, respectively). This method of free recall enables the researcher to assay the underlying content of memory using a model-based approach (e.g., Anderson & Awh, 2012; Bays, Catalao, & Husain, 2009; Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011; Poliakov, Stokes, Woolrich, Mantin, & Astle, 2014). In particular, it is possible to estimate the proportion of trials upon which participants correctly retain a representation of a target item (and the precision with which they do so), those upon which they incorrectly report a non-target representation, and the proportion of trials upon which participants simply guess (Fig. 1). In short, unlike the methods typically used in developmental populations, this continuous report methodology, alongside mathematical modeling, provides the researcher with an assay of the underlying content of memory. However, to our knowledge, this has only once been applied to separate these processes in childhood (Burnett Heyes et al., 2012). Children, aged between 7 and 13, were presented with either one or three oriented bars in sequence, and after a brief delay were asked to report the orientation of one of those bars. For the first time, the authors were able to show that between the ages of 7 and 13 the precision with which items could be retained showed a linear increase. This led the authors to suggest that one possible mechanism for developmental improvements in visuospatial STM is the gradual improvement in maintenance accuracy, rather than an increase in the discrete number of items that can be maintained per se.

In this study we used a novel paradigm in which children were presented with to-be-remembered bars in sequence. Following a brief delay the children had to attempt to report either the exact orientation, or the mirror image of the orientation, of one of the bars in the sequence. That is, on some trials children simply maintained and reported orientations, whereas on other trials there was an additional requirement of online manipulation (mental rotation). A trial schematic can be seen in Fig. 2. Our first aim was to explore the impact of this concurrent processing on subjects' memory of the target orientation, and then to use the probabilistic modeling approach described above to explore in more detail the impact of the mental rotation on memory performance. Our second aim was to explore developmental differences on both trial types, again using the modeling to separate the potential contributions to performance improvements with age. Finally, we also collected performance measures on standardized visuospatial STM and WM tasks. This allowed use to explore how performance on our continuous performance task, and the underlying parameters produced by the modeling, would vary across children according to individual differences in visuo-spatial STM and WM capacity.
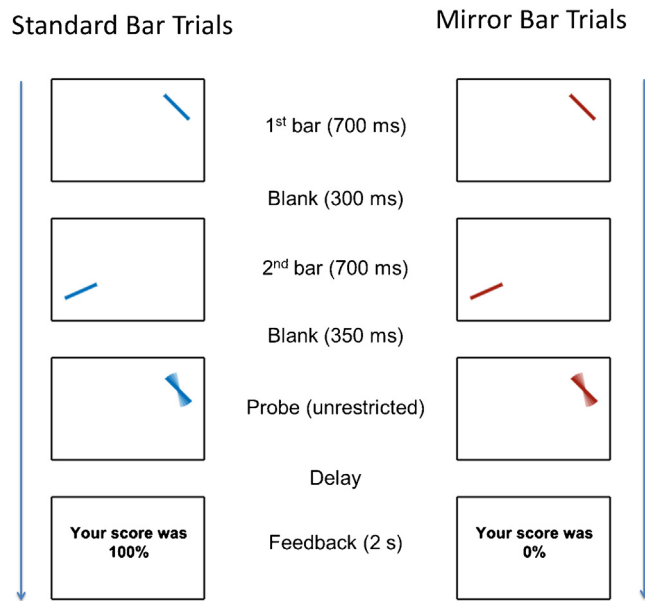
Standard Bar Trials   Mirror Bar Trials

Fig. 2. Trial order schematic for our Standard and Mirror Bar conditions.

2. Method

2.1. Participants

Twenty children (9 boys) between 7 and 9 years old (*M* = 8.8 years old, SD = 0.10; "Younger Children" henceforth) and 20 children (14 boys) between 10 and 12 years old (*M* = 11.1 years old, SD = 0.20; "Older Children" henceforth) participated in the study. For our novel experimental task we also collected data from 21 adults (9 males) between 21 and 39 (*M* = 25.62 years old, SD = 0.91, "Adults" henceforth) to provide a measure of optimal performance. Children were recruited from local primary schools. We used a composite score of our standardized spatial STM and WM tasks as an index of the children's general cognitive ability (Alloway, Gathercole, Kirkwood, & Elliott, 2009; Conway et al., 2005; Fukuda, Vogel, Mayr, & Awh, 2010; Johnson et al., 2013). Our sample of children had a mean age-standardized composite score of 112.56 (±12.70 SD). All participants had normal or corrected-to-normal vision. Prior to testing, ethical approval from the University of Cambridge Research Ethics Board was secured. Adult participants and parents/guardians of children provided written informed consent.

2.2. Cognitive measures

2.2.1. Continuous report task

A trial schematic of the paradigm can be seen in Fig. 2. Participants viewed two oriented target bars (approx. 2° × 0.3° of visual angle) that appeared sequentially at one of four possible locations for 700 ms each, separated by a 300 ms blank inter-stimulus interval. This timing was chosen in order to allow sufficient time for children of all ages to encode the visual information (Cowan, AuBuchon, Gilchrist, Ricker, & Saults, 2011; Cowan, Morey, AuBuchon, Zwilling, & Gilchrist, 2010). After the second blank interval of 300 ms and an additional delay of 50 ms, a white probe bar appeared at the same location of one of the target bars. Participants had to adjust the orientation of the probe bar so that it matched their memory of either the orientation, or the mirror image, of the original target bar. In order to achieve this, participants used two keyboard buttons that would rotate the probe bar either clockwise or anti-clockwise. All bars in the sequence were probed with equal probability, and there was no time limit on the response. As soon as participants were satisfied with their response, they pressed a keyboard button to move to the next trial. Following this they were given feedback from 0 to 100%, according to how accurately they adjusted the probe to the target bar, presented for 1000 ms. There was then a break of 500 ms before the start of the next trial. Additionally, at the end of each block participants received feedback concerning their average performance in that particular block ranging from 0 to 100%. The background was kept black during the whole task. Bar orientation (targets, probe) was independently randomized across π rad on each trial.

The task comprised two conditions, the Standard Bars condition and the Mirror Bars condition presented in an alternating blocked design. In the Standard blocks, participants had to match the probe bar to the exact orientation of the target bar whilst in the Mirror blocks, to the mirror image (along the vertical meridian) of the target bar. Mental rotation is a skill that can be seen in children as early as 5 months (Moore & Johnson, 2008), develops considerably in the pre-school years (Kail, Pellegrino, & Carter, 1980), and continues to develop well into adolescence (Kail, 1985). In visuo-spatial WM paradigms,

mental rotation is often used to as the current processing element, in order to produce a complex span task (e.g., Gathercole, Pickering, Ambridge, & Wearing, 2004). This is also true of the standardized WM task that we implemented here (details below). In order to help participants with the different block requirements, target bars were colored according to condition: in the Standard blocks they were blue, whilst in the Mirror blocks they were red.

The task started with a short practice: 5 Standard and 5 Mirror trials. Subjects completed a total of 10 blocks, with each block consisting of 15 trials; overall 75 trials of the Standard and 75 trials of the Mirror conditions. After each block individuals were encouraged to take a short break.

2.2.2. Automated Working Memory Assessment (AWMA; Alloway, 2007).

In addition to our bespoke continuous report task, we also administered two subtests of the AWMA. These were the Dot Matrix task and the Spatial Recall task, and are considered to tap spatial STM and spatial WM, respectively.

In the Dot Matrix task, participants view a 4×4 matrix and sequentially presented red dots. They recall where dots appeared in the matrix in exactly the same order as they saw them by pointing to the computer screen. In the Spatial Recall task, participants view two identical shapes, one of which is the mirror image of the other on half of all trials. They report whether the two shapes are a mirror image of each other, thus performing a mental rotation on the second shape. At the same time, the second shape is paired with a red dot whose serial positions participants recall at the end of the sequence. The task therefore requires the child to retain the ordered locations of a dot while mentally rotating objects. For both tasks the number of to-be-remembered items increases until below 2/3 criterion. Raw scores and standardized scores were computed for both subtests.

Children completed all the above-mentioned tasks in a relatively quiet place in school, during school hours. The sequence in which these tests were given was counterbalanced to prevent order effects. We included additional measures of literacy and numeracy, which are not discussed here. The session for children typically lasted around an hour. If that limit was exceeded, the child completed testing on another occasion in order to avoid fatigue effects. Adults completed only the continuous report task, which lasted approximately half an hour, always in a quiet room. Children and adults received verbal instructions about the continuous report task. Particular emphasis was placed on making sure that the children understood what a mirror image is. For that, they practiced drawing mirror images of bars on a tablet.

2.3. Analytic approach to the continuous report data

Participants' continuous responses were analyzed with respect to the probe bar orientation, and how different this was to the actual target bar orientation (or its mirror, depending upon the condition). This was achieved by calculating the overall error in response angle. The responses were coded in the range of 0–180° (the range of unique angles in our paradigm). From this we calculated the overall precision (1/SD) with which subjects responded. This value corresponds directly to the overall spread of participant's errors. Subsequently we employed a modeling approach to explore subjects' responses more closely, with the model being applied to each subject on an individual basis.

2.3.1. Raw score (1/SD)

For each trial the angular deviation between the response orientation and the actual orientation of the target bar was calculated (i.e., the angular error). Precision was subsequently calculated for each participant, as the reciprocal of the standard deviation of error across trials (1/SD). Considering that the parameter space for orientation is circular, Fisher's definition of SD for circular data was used, adjusting for random responses (Fisher, 1995). More detailed descriptions of this method can be found elsewhere (Bays et al., 2009, 2011; Gorgoraptis, Catalao, Bays, & Husain, 2011). This procedure leads to a generic measure of the resolution with which the target orientation is reproduced, called "precision"; i.e., how concentrated the distribution of responses was.

2.3.2. Applying a standard mixture model

The raw score from the continuous report task gives an indication of the overall variability in responses. In order to further characterize the responses produced by our participants, a standard mixture model was applied to the data collected using either the standard or mirror trials. This model has been previously used in adults (Bays et al., 2009; Zhang & Luck, 2008) and children (Burnett Heyes et al., 2012). The distribution of responses (using a mixture model from Bays et al., 2009) was described using four parameters: (i) κ (Kappa), a concentration parameter encapsulating Gaussian variability in memory for target orientations. In other words, Kappa represents how precisely individuals remembered the target orientation when they responded with the orientation of the correct item. (ii) p(T), the probability that participants responded with the target orientation. (iii) p(NT), the probability that participants erroneously responded with the non-target orientation; on some occasions individuals might confuse which of the two bars they need to report, and may report the wrong one. This is considered to constitute a misbinding error (Bays et al., 2009). (iv) p(U) represents the probability that participants made a uniform error (a random guess).
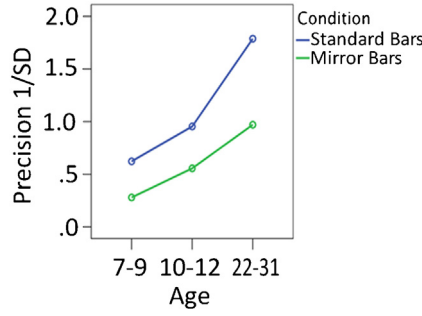
Fig. 3. Mean precision (1/SD) values for both trial trypes across all three groups.

The mixture model can be described by the following equation (Bays et al., 2009, 2011):

$$p\left(\hat{\theta}\right) = \alpha\phi k\left(\hat{\theta} - \theta\right) + \beta\frac{1}{m}\sum_{i}^{m}\phi_{\kappa}\left(\hat{\theta} - \varphi_{i}\right) - \gamma\frac{1}{2\pi}.$$

with θ corresponding to the correct orientation of the target item and $\hat{\theta}$ being the orientation reported by the subject. $\phi\kappa$ corresponds to the Von Mises distribution—a Gaussian distribution equivalent for a circular response space, with a mean of 0 and concentration κ. α is the probability of reporting the target item; β is the probability of reporting a non-target item, with m corresponding to the number of potential non-target items. The probability of responding randomly can be calculated using $\gamma = 1 - \alpha - \beta$. Further details on the implementation of this mixture model can be found at: http://www.paulbays.com/code/JV10/.

3. Results

The results are organised in the following way: we first analyze the precision values (the overall spread in responses), testing for differences between our conditions, and across the three groups; secondly we apply mixture modeling and explore the extent to which model parameters are influenced by condition (Standard versus Mirror trials) and age; finally we explore the relationship between the model parameters and standardized STM and WM measures.

3.1. Analysis of precision values

To test for differences between our conditions, and across age groups we submitted our precision values (1/SD) to a 3 (Age: Younger Children, Older Children and Adults) by 2 (Condition: Standard, Mirror) ANOVA. Before doing this, we used our age-standardized memory scores to check that the two groups of children were of equivalent ability for their respective ages. Independent t-tests revealed that a composite score of standardized measures of visuospatial memory (Dot Matrix and Spatial Recall) did not differ significantly between the two groups ($t(38) = 0.425$, $p = 0.673$). Subsequently we can be confident that any differences between our two groups of children can be attributed to development, rather than to one group of children being of a relatively higher ability level for their age.

The mean precision values for both conditions, across the three age groups, can be seen in Fig. 3. The 3 × 2 ANOVA revealed a significant main effect of Age ($F(2,58) = 30.993$, partial $\eta^2 = 0.517$, $p < .001$). Tukey post-hoc tests revealed that this was because the Older Children were significantly more precise than the Younger Children ($p = 0.046$), and the Adults were more precise than both Younger and Older Children ($p < 0.001$ and $p < 0.001$ respectively), when collapsing across conditions. In short, overall precision increased with age. There was also a significant main effect of condition, with responses on Standard trials being significantly more precise than on Mirror trials ($F(1,58) = 88.531$, partial $\eta^2 = 0.604$, $p < 0.001$). The interaction between the linear effect of age and condition was significant ($F(2,58) = 4.475$, partial $\eta^2 = 0.134$, $p = 0.016$). This is because the difference between the conditions was significantly or marginally significantly greater for Adults, relative to the Younger and Older groups of children ($p = 0.020$ and $p = 0.059$ respectively; Tukey test). The difference between the Standard and Mirror conditions was not significantly different between Younger and Older Children ($p = 0.900$; Tukey test). In short, as subjects became older they became more precise, and the precision difference between the two conditions significantly increased into adulthood.

3.2. Probablistic modeling of continuous responses

The modeling parameters for one child on Standard trials and for four children on Mirror trials were >2.5 SD beyond the mean for their respective age groups, and these children were thus removed from the further analysis. The modeling parameters are not independent of one another and therefore cannot be submitted to an ANOVA together. Instead we

addressed two separate questions with these data. Firstly, we used the children's data to explore the effect of condition (Standard versus Mirror Trials) on the various parameters from the model. Secondly, taking each condition separately, we explored developmental differences in each model parameter.

3.2.1. Effect of mental rotation on model parameters

In this analysis we just considered the children's data, since we are collapsing across age and it may not be appropriate to collapse across very large age differences. Our model fitting procedure revealed a number of significant differences between the Standard and Mirror conditions. There were significantly more concentrated responses to targets [Kappa ($t(34) = 4.772$, $p < 0.001$)], a significantly higher proportion of correct responses to the target [pT ($t(34) = 3.343$, $p = 0.002$)], and significantly fewer guesses [pU ($t(34) = -2.491$, $p = 0.018$)] on Standard relative to Mirror trials. However, we could find no difference in the level of erroneous non-target responses [pNT ($t(34) = -0.572$, $p > 0.05$)]. These data can be seen in Fig. 4A. To highlight the differences in concentration coefficient, the Kappa differences across the two conditions are shown in Fig. 4B as probability density functions. This illustrates the greater concentration of responses around the 0 point (which would be a perfect response) for the Standard, relative to the Mirror, condition. In short, the concurrent processing demand of the mental rotation made the children significantly more likely to guess, and when they did remember the target item, they were less precise.

3.2.2. Developmental differences in model parameters for Standard trials

The model parameters for the Standard trials can be seen in Fig. 4C. One-way ANOVAs indicated that the developmental differences in precision (1/SD) we observed in our previous analysis could be attributed to more precise responses to targets [Kappa ($F(2,59) = 8.137$, $p = 0.001$)], a significantly higher number of correct responses to the target [pT ($F(2,59) = 11.887$, $p < 0.001$)], and significantly fewer guesses [pU ($F(2,59) = 14.540$, $p < 0.001$)] with age. However, we could find no difference in the level of erroneous non-target responses [pNT ($F(2,59) = 1.140$, $p = 0.327$)]. On some occasions the assumption of homogeneity of variances was not met but the robustness of the analysis was confirmed using non-parametric tests (Independent-samples Kruskal–Wallis) that yielded the same results.

We subsequently explored pair-wise differences using Tukey post-hoc comparisons, in order to establish where the group differences were most apparent. This revealed that there were no significant differences in how precisely the two groups of children could report the Standard orientation [Kappa, $p = 0.493$], whereas Adults reported the Standard orientation more accurately compared to both Younger [$p = 0.001$] and Older Children [$p = 0.020$]. The Younger Children were significantly less likely to report the target compared to Older Children [$p(T)$, $p = 0.012$] and Adults ($p < 0.001$). There were no significant differences in the probability of reporting the target between Older Children and Adults [$p(T)$, $p = 0.162$]. The results were similar concerning random guesses: The Younger Children were significantly more likely to make a random guess compared to older children [$p(U)$, $p = 0.006$] and adults [$p < 0.001$]. There were no significant differences in the probability of random guesses between Older Children and Adults [$p(U)$, $p = 0.094$]. This implies that the improvement in overall precision that children show with age stems from an improvement in the discrete probability that the item is represented in memory, rather than an improvement in the quality with which that representation can be maintained or reported. By contrast developmental improvements into adulthood were also apparent in changes in Kappa—the precision with which targets can be retrieved.

An alternative way of performing this analysis is to treat all of the children as a single group, and explore the effect of age as a continuous variable in a linear regression analysis. This confirmed the findings of the previous analysis: the guessing rate significantly reduced [$p(U)$: standardized beta coefficient = -0.438, $t = -2.900$, $p = 0.006$] and the proportion of targets reported significantly increased with age [$p(T)$: standardized beta coefficient = 0.485, $t = 3.435$, $p = 0.002$], however there was no significant age effect on the precision with which targets were reported [Kappa: standardized beta coefficient = 0.078, $t = 0.474$, $p = 0.638$], or on the likelihood of misreporting the non-target item [$p(NT)$: standardized beta coefficient = -0.194, $t = -1.173$, $p = 0.248$].

3.2.3. Developmental differences in model parameters for Mirror trials

We analyzed the Mirror trials in exactly the same way as we had for the Standard trials. The model parameters for the Mirror trials can be seen in Fig. 4D. One-way ANOVAs indicated that the development differences in precision we had previously observed could be attributed to more precise responses to targets [Kappa ($F(2,56) = 15.334$, $p < 0.001$)], a significantly higher number of correct responses to the target [pT ($F(2,56) = 9.138$, $p < .001$)], and significantly fewer guesses [pU ($F(2,56) = 10.023$, $p < 0.001$)]. However, we could find no difference in the proportion of erroneous non-target responses [pNT ($F(2,56) = 0.467$, $p = 0.629$)]. On some occasions the assumption of homogeneity of variances was not met but the robustness of the analysis was confirmed using non-parametric tests (Independent-samples Kruskal–Wallis) that yielded the same results.

Again, we used Tukey post-hoc comparisons to explore where the group differences were most apparent. There were no significant differences in how precisely the two groups of children could report the Mirror orientation [Kappa, $p = 0.998$]. However, the Adults were significantly more precise in reporting the Mirror orientation compared to Younger Children ($p < 0.001$) and Older Children ($p < 0.001$). The Younger Children were significantly less likely to report the target compared to Adults [$p(T)$, $p < 0.001$] but not compared to Older Children [$p = 0.129$]. There were no significant differences in the probability of reporting the target between Older children and Adults [$p(T)$, $p = 0.077$]. There were no significant differences
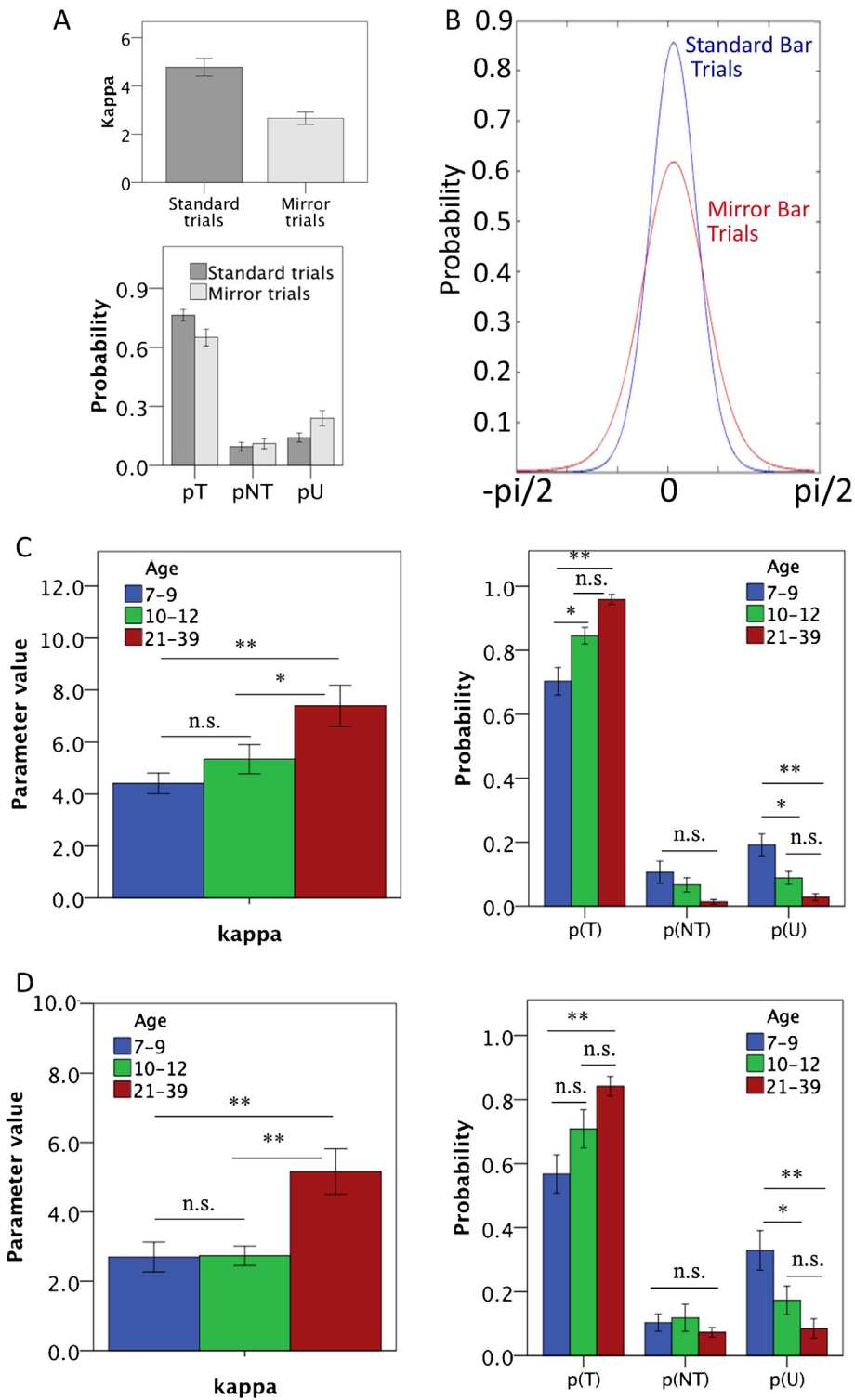
Fig. 4. (A) Parameters from the Mixture Model for Standard and Mirror trials. The top panel shows the concentration of responses to the target item (Kappa). The bottom panel shows the proportion of trials upon which the target was correctly reported (*p*T), a non-target item was reported (*p*NT) and the proportion of trials upon which the child guessed (*p*U). (B) Probability Density Functions for correct target responses in our two conditions. (C) The model parameters for all three groups on the Standard trials. The left hand panel shows the concentration of target responses (Kappa). The right hand panel shows the proportions of trials that correspond to correct target responses (*p*T), non-target responses (*p*NT) and uniform guesses (*p*U). (D) Shows the same parameters as for C, but for Mirror trials.

Table 1
Pearson correlation values controlling for age.

Control variable	Measure	1	2	3	4	5	6	7	8	9	10	11
Age (months)	1. Dot Matrix											
	2. Spatial recall	0.46[*]										
	3. Standard Bars 1/SD	0.51[**]	0.61[***]									
	4. Standard Bars Kappa	−0.01	0.01	−0.08								
	5. Standard Bars pT	0.44[**]	0.46[**]	0.81[***]	−0.48[**]							
	6. Standard Bars pNT	−0.1	−0.14	−0.43[**]	−0.20	−0.44[**]						
	7. Standard Bars pU	−0.39[*]	−0.39[*]	−0.53[***]	0.67[***]	−0.73[***]	−0.30					
	8. Mirror Bars 1/SD	0.25	0.41[*]	0.67[***]	0.05	0.40[*]	−0.26	−0.23				
	9. Mirror Bars Kappa	0.08	0.01	0.22	0.11	0.13	−0.13	−0.04	0.28			
	10. Mirror Bars pT	0.31	0.45[*]	0.55[***]	−0.11	0.47[**]	−0.27	−0.29	0.65[***]	−0.23		
	11. Mirror Bars pNT	−0.34[*]	−0.21	−0.33[*]	0.06	−0.37[*]	0.38[*]	0.10	−0.27	−0.19	−0.35[*]	
	12. Mirros Bars pU	−0.11	−0.33[*]	−0.37[*]	0.07	−0.26	0.04	0.24	−0.50[**]	0.35[*]	−0.81[**]	−0.26

[*] $p < 0.05$.
[**] $p < 0.01$.
[***] $p < 0.001$.

in the probability of making random guesses between Older children and Adults [p(U), $p = 0.158$], whereas the Younger Children were significantly more likely to make a random guess compared to Adults [p(U), $p < 0.001$] and Older Children [p(U), $p = 0.040$]. In summary, the pattern of findings is very similar to that for the Standard trials: developmental differences between the two groups of children are restricted to the discrete probability of an item being represented in memory (i.e., the guessing rate), and not the precision with which children can remember the items. By contrast developmental differences into adulthood are associated with both changes in this discrete probability and improved precision.

An alternative way to perform this analysis is to treat all of the children as a single group, and explore the impact of age as a continuous regressor on the various parameters of the model. This approach broadly confirmed the result of the previous analysis: as children got older they were significantly more likely to report the target bar [p(T): standardized beta coefficient = 0.358, $t = 2.250$, $p = 0.031$] and significantly less likely to guess [p(U): standardized beta coefficient = −0.344, $t = −2.089$, $p = 0.045$]. But older children were no more precise [Kappa: standardized beta coefficient = −0.078, $t = −0.451$, $p = 0.655$] or less likely to mistakenly report the non-target item [p(NT): standardized beta coefficient = −0.071, $t = −0.406$, $p = 0.687$].

3.3. Relationships across memory measures

We also explored how a child's performance on widely-used standardized STM and WM capacity measures would relate to these modeling parameters. The first step we took in doing this was to perform a set of pair-wise correlations.

3.3.1. Correlations

A correlational analysis was performed between the measures from our continuous report task and measures from the AWMA battery. Table 1 shows partial correlations controlling for age. These age-controlled correlations highlight a number of noteworthy relationships: (i) our two standardized measures are significantly correlated with one another. (ii) Both standardized measures are significantly correlated with precision (1/SD) on Standard bar trials, although their relationship with precision on Mirror bar trials is more modest. Indeed, only the Spatial Span task significantly correlates with precision on the Mirror bar trials. This overall reduction in the strength of relationship may reflect the more variable performance of children on the more difficult Mirror bar trials, and that the best relationship is with the Spatial Span task may reflect the common mental rotation component to both the Spatial Span task and the Mirror bar condition. (iii) The parameters from the model are closely related to one another. This is necessarily the case, since these parameters will trade-off against one another in the fitting of the mixture model. For example, as pT increases, so pU and pNT will decrease—if the proportion of trials classed as correct recalls increases then the proportion of trials classed as guesses or binding errors will necessarily fall. In short, the proportion parameters (pT, pU and pNT) will all be negatively correlated with one another. This can be seen in Table 1 for both the Standard and Mirror trials. From Table 1 we can also see that Kappa is positively correlated with pU and negatively correlated with pT in both the Standard and Mirror conditions, although this relationship is most robust in the Standard condition. We can think of two potential reasons for this. Firstly, the model itself may produce this. Responses that are relatively far from the correct orientation could either be characterized as inprecise correct responses or guesses. Depending upon how the mixture model allocates these responses, in order to produce the best fit, could influence Kappa. For example, if those responses are classed as correct then pT will go up, but the overall precision of the correct response distribution (Kappa) will drop, since more inprecise responses are included. Secondly, children may adopt a strategy that produces a trade-off. By being very conservative they may only report targets when they are very sure of their correct orientation, and if they are unsure then they may just guess. This would result in a small proportion of correct, but highly precise responses, and a large number of guesses. (iv) The modeling parameters are significantly related to standardized

Table 2
Regression tables. The top half shows the regression Table for Dot Matrix performance as the outcome variable. The bottom half shows the same but with Spatial Span as the outcome variable.

	Model parameters				Coefficients		
	Adjusted R^2	R^2 change	F change	p	Standardized β	t	p
Dot matrix (raw)							
Step 1	0.017	0.043	5.68	0.022			
Age					0.365	2.383	0.022
Step 2	0.367	0.3	6.013	0.002			
Kappa					0.43	3.217	0.003
p(NT)					−0.182	−1.307	0.2
p(U)					−0.465	−3.021	0.005
Spatial recall (raw)							
Step 1	0.163	0.185	8.387	0.006			
Age					0.43	2.896	0.006
Step 2	0.413	0.29	6.251	0.002			
Kappa					0.376	2.921	0.006
p(NT)					−0.222	−1.65	1.08
p(U)					−0.497	−3.346	0.002

measures. Children who have higher scores on the standardized measures tend to have higher pT and lower pU parameters, although for the Mirror trials this relationship only holds significantly for the Spatial Span measure.

These correlations informed our subsequent linear regression analyses, in which we explored the extent to which the modelling parameters would predict a child's performance on the standardized STM and WM measures.

3.3.2. Linear regressions

We were keen to explore the extent to which variability in memory capacity across children, as established using the standardized measures from the AWMA, could be explained by the various parameters from our mixture model. To do this we performed two regressions. In the first our outcome measure was capacity on the Dot-Matrix task. In our first step we included Age in months. In our second step we included the Kappa, p(NT) and p(U) parameters. We did not include p(T) because, as can be seen in Table 1, this is very closely related to p(U) and Kappa, and including it introduces a multicolinearity problem into the regression. We found no evidence for multicollinearity between the rest of the predictors (mean VIF < 2). We then performed the same regression, but using Spatial Recall as our outcome measure. For both of these regressions we used the parameters from the Standard trials. The correlational analysis showed that the pattern of relationships between the model parameters and standardized measures was very similar across both conditions, but that these relationships were most robust and consistent in the simpler Standard bar trials. Furthermore, we had reliable parameter estimates for all bar one child for the Standard bar condition, whereas four children were removed from the Mirror bar condition as outliers. In all of our regressions, rather than using the age-standardized scores from the AWMA, we used raw scores. This is because the age-standardized scores are age standardized in whole year sections in the AWMA, and using standardized scores can alter the distribution of scores. Instead, age is controlled for in our analyses by including age in months as the first entry in each regression.

The results of these regressions can be seen in Table 2. Once age was controlled for, a child's score on the Dot-Matrix task could be significantly explained by the precision of their target responses and the proportion of guesses. The same was true of the Spatial Recall task. In short, whilst developmental differences memory performance in childhood might be restricted to the discrete probability of an item being recalled, individual differences between children are apparent in both this probability and in the precision with which targets can be recalled.

4. Discussion

Children's ability to hold in mind visuospatial information is strictly limited, it develops gradually, and individual differences across children are significantly related to their level of educational attainment (for a review see Cowan, 2014). Relying upon standardized measures alone does not currently allow us to address a number of important questions: (i) which aspects of memory performance are impacted upon by a concurrent processing demand?; (ii) which aspects of our memory improve as we develop?; and (iii) which component processes correspond most closely to individual differences in memory capacity across children? Our continuous report paradigm required children to hold in mind the orientation of bars, and then report a cued bar at the end of the trial. We could use the responses of our participants to measure how precisely they could report the target. We then applied a standard mixture model to derive four components from the response data—the concentration coefficient with which targets were reported (Kappa), the proportion of trials upon which a target was reported (pT), the proportion of trials upon which children erroneously reported the uncued item (pNT), and the proportion of trials upon which children guessed (pU).

4.1. Concurrent processing reduces the quality of item representation and reduces the likelihood of it being represented in memory

Working memory refers to the concurrent maintenance and processing of information for brief periods of time. In visuo-spatial WM tasks, mental rotation is often used in order to provide this additional processing demand. A child's WM capacity is considered an especially good predictor of their level of educational attainment (Engel de Abreu, Conway, & Gathercole, 2010; Engel de Abreu, Gathercole, & Martin, 2011; Gathercole, Alloway, Willis, & Adams, 2006), however, the impact of this concurrent processing on memory is unclear (Aben, Stapert, & Blokland, 2012; Conway et al., 2005; Cowan, 2008). One possibility is that the concurrent processing reduces the representational quality of items in short-term storage. Alternatively, it could increase the probability of items being lost from memory (forcing children to guess) or it could increase the probability of children confusing the order of stimuli (increasing the likelihood of a misbinding error).

In our data concurrent processing acted to both reduce the quality of item representation, and reduce the likelihood of the target item being represented in memory. When children had to retain the orientation of the bar whilst performing a mental rotation they were significantly less precise in their memory for the target, and the discrete probability of them guessing also increased. In short, the additional processing demand of performing the mental rotation had a non-specific influence on short-term storage, with the only parameter unaffected being the misbinding probability. However, it is important to note that a recent factor analysis of ten experiments using continuous report in adults has indicated that misbinding errors explain only a very small proportion of responses (van den Berg, Awh, & Ma, 2014). As such, it may be the case that we have insufficient power with our design to detect misbinding error differences with and without concurrent processing. It is also important to note that our task only uses two items in sequence, which may also make misbinding errors unlikely. If children were presented with longer sequences of to-be-remembered orientations, akin to the longer sequences of items used in standardized STM and WM assessments, then misbinding errors may indeed be sensitive to concurrent processing. This notwithstanding, in so far as we can discriminate these processes in our data, the additional demand of performing a mental rotation negatively influenced the children's ability to report the target by both increasing the guessing rate and reducing the precision of the target representation.

4.2. With age children are more likely to correctly report the target

We observed substantial developmental improvements in overall precision on our continuous report task; Adults were significantly more precise than Older Children, who were in turn more precise than Younger Children. Applying the probabilistic model revealed that adults were superior on all parameters with the exception of pNT (which represents misbinding errors). Previous studies have suggested that memory improvement with development from childhood to adulthood entails a decrease in the number of misbinding errors (Cowan et al., 2006; Oakes, Ross-Sheehy, & Luck, 2006). Whilst our findings might appear to contradict this, it is again important to consider that misbinding errors (pNT) are not frequent in continuous report tasks and thus we might not have enough power to detect the relevant developmental differences. Our data do however clearly demonstrate that the superior performance of adults relative to children stems from more precise representations of the target item, in combination with an increase in the discrete probability of the target item being reported.

The differences between the two groups of children were more specific. Older Children were significantly more likely to report the target, and significantly less likely to guess, relative to the Younger Children. However, their responses to the target were no more concentrated. In short, developmental improvements on our continuous report task during childhood were underpinned by improvements in the discrete probability that the target was represented in memory, rather than in improvements in the quality of that representation. This overall pattern of effects was very similar across both the Standard and Mirror conditions, and can perhaps be most clearly seen in the linear regression analyses that use age as a continuous variable. It is important to note that the modeling is still valid across children of different ages, even though the overall guessing rate is different. As the number of target responses drops (i.e., as guessing increases) it could make the Kappa estimates more variable, but it would not result in us systematically over or underestimating Kappa. As it happens in our data the variability of Kappa estimates actually appear to be very similar for both groups of children, suggesting that even though the younger group have fewer target responses, there are still enough to accurately assay the precision of those responses.

This result seems to contrast with the only previous demonstration of this technique with developmental populations (Burnett Heyes et al., 2012). However, it is important to note that there are a number of differences across these two studies. The previous experiment found that developmental differences in modeling parameters were observed only between early adolescents (13-year-olds) and children (9–12 years old), and that these were specific to precision (Kappa). By contrast, in our study developmental differences were found between children aged 8 and 11. It is possible that the quality of the memory representation does increase from late childhood into early adolescence, but that we did not capture this with our age groups. We certainly see precision differences between older children and adults, so this is entirely plausible. However, in their study no developmental improvements were found in the discrete probability that the target was represented in memory, which is in contrast with our findings. The task used by Burnett-Hayes and colleagues was arguably more difficult than the one employed here: children were asked to keep in mind three bars rather than two. This overall increased difficulty could introduce variability in the parameter estimates for some of the younger children. In short, we do not think that our

findings are necessarily discrepant with those of Burnett-Hayes and Colleagues – differences in participant age and in overall task difficulty could have a very significant impact upon which parameters show the most marked improvement with age.

In summary, in our data developmental improvements from childhood to adulthood were underpinned by both improvements in the discrete probability that an item is represented, as well as by improvements in precision. By contrast, developmental improvements within childhood were underpinned by improvements in just the likelihood of the item being reported, and not by any change in precision per se.

4.3. Individual differences in visuospatial memory capacity correspond to modeling parameters

Performing poorly on standardized measures of STM, and in particular WM, is a very significant risk factor for both educational underachievement and neurodevelopmental disorder (Alloway, Gathercole et al., 2009). It is therefore important that we consider what underlying processes might contribute to individual differences on these measures. Using a linear regression we were able to show that significant variance in performance on a standardized measure of visuospatial STM (the Dot Matrix task) and WM (the Spatial Recall task) could be explained by two model parameters: how concentrated their responses were on the continuous report task (increased Kappa) and how likely they were to guess. This is an interesting contrast to our developmental effects—the Older Children produced fewer guesses, but were no more precise than the Younger Children. One possibility is that the parameters that drive developmental differences can be distinguished from those that drive individual differences (see also, Astle et al., 2014). That is, what makes children particularly good for their respective age need not be the same mechanism that drives the large improvements in ability that occur with age. This distinction also mirrors neuroimaging findings, which show that complementary neural systems support visuo-spatial memory maintenance in adulthood. Specifically, the activity of one system centered on the inferior intra-parietal sculcus reflects the number of items being retained, regardless of the complexity of the items; by contrast another system, centered on the superior intra-pareital cortex and lateral occipital cortex, reflects a variable set of attended objects, depending upon their complexity (Xu & Chun, 2006). It is possible that the different parameters produced by mixture models are sensitive to these complementary systems, with the first system reflecting the number of items that can be retained, and the second reflecting the graded precision with which items can be retained. In our data, individual differences in 'capacity' are reflected in both of these processes, whereas developmental differences are only apparent in the former.

5. Conclusion

Performance on standardized measures of STM and WM have enabled to us to chart developmental and individual differences in visuospatial memory in childhood. Here we show that continuous report measures provide important additional information about these processes in childhood. An additional processing demand acts to both reduce the precision of children's responses, and reduce the likelihood of the memoranda being represented in memory. Within childhood developmental improvements in visuospatial memory performance are most apparent in a reduction in the discrete probability of children guessing (i.e., they are more likely to have retained some representation of the target), but Older Children are no more precise than Younger Children. Finally, a child's performance on standardized measures of visuospatial STM and WM can be predicted by both their precision and guessing rate on the continuous report paradigm.

References

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in Psychology*, 3, 301. http://dx.doi.org/10.3389/fpsyg.2012.00301

Alloway, T. P. (2007). *Automated working memory assessment (AWMA).* London: Pearson Assessment.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development*, 80(2), 606–621. http://dx.doi.org/10.1111/j.1467-8624.2009.01282.x

Anderson, D. E., & Awh, E. (2012). The plateau in mnemonic resolution across large set sizes indicates discrete resource limits in visual working memory. *Attention, Perception & Psychophysics*, 74(5), 891–910. http://dx.doi.org/10.3758/s13414-012-0292-1

Astle, D. E., Harvey, H., Stokes, M., Mohseni, H., Nobre, A. C., & Scerif, G. (2014). Distinct neural mechanisms of individual and developmental differences in VSTM capacity. *Developmental Psychobiology*, 56(4), 601–610. http://dx.doi.org/10.1002/dev.21126

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10), 7. http://dx.doi.org/10.1167/9.10.7

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, 11(10) http://dx.doi.org/10.1167/11.10.6

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321(5890), 851–854. http://dx.doi.org/10.1126/science.1158023

Burnett Heyes, S., Zokaei, N., van der Staaij, I., Bays, P. M., & Husain, M. (2012). Development of visual working memory precision in childhood. *Developmental Science*, 15(4), 528–539. http://dx.doi.org/10.1111/j.1467-7687.2012.01148.x

Conway, A. R. A., 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*(5), 769–786.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. http://dx.doi.org/10.1016/S0079-6123(07)00020-9

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223. http://dx.doi.org/10.1007/s10648-013-9246-y

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*(5), 1066–1074. http://dx.doi.org/10.1111/j.1467-7687.2011.01060.x

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*(1), 120–133. http://dx.doi.org/10.1111/j.1467-7687.2009.00864.x

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: when is feature binding difficult? *Developmental Psychology*, *42*(6), 1089–1102. http://dx.doi.org/10.1037/0012-1649.42.6.1089

Engel de Abreu, P. M. J., Conway, A. R. A., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, *38*(6), 552–561. http://dx.doi.org/10.1016/j.intell.2010.07.003

Engel de Abreu, P. M. J., Gathercole, S. E., & Martin, R. (2011). Disentangling the relationship between working memory and language: the roles of short-term storage and cognitive control. *Learning and Individual Differences*, *21*(5), 569–574. http://dx.doi.org/10.1016/j.lindif.2011.06.002

Fisher, N. I. (1995). *Statistical analysis of circular data.* Cambridge: University Press.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679. http://dx.doi.org/10.3758/17.5.673

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, *93*(3), 265–281. http://dx.doi.org/10.1016/j.jecp.2005.08.003

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*(2), 177–190. http://dx.doi.org/10.1037/0012-1649.40.2.177

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience*, *31*(23), 8502–8511. http://dx.doi.org/10.1523/JNEUROSCI.0208-11.2011

Johnson, M. K., McMahon, R. P., Robinson, B. M., Harvey, A. N., Hahn, B., Leonard, C. J., . . . & Gold, J. M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology*, *27*(2), 220–229. http://dx.doi.org/10.1037/a0032060

Kail, R. (1985). Development of mental rotation: a speed-accuracy study. *Journal of Experimental Child Psychology, 40*(1), 181–192.

Kail, R., Pellegrino, J., & Carter, P. (1980). Developmental changes in mental rotation. *Journal of Experimental Child Psychology, 29*(1), 102–116.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: a sex difference. *Psychological Science*, *19*(11), 1063–1066. http://dx.doi.org/10.1111/j.1467-9280.2008.02200.x

Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science*, *17*(9), 781–787. http://dx.doi.org/10.1111/j.1467-9280.2006.01782.x

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, *32*, 301–345. http://dx.doi.org/10.1016/0001-6918(70)90108-3

Poliakov, E., Stokes, M. G., Woolrich, M. W., Mantini, D., & Astle, D. E. (2014). Modulation of alpha power at encoding and retrieval tracks the precision of visual short-term memory. *Journal of Neurophysiology, 112*(11), 2939–2945.

Van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. http://dx.doi.org/10.1037/a0035234

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135. http://dx.doi.org/10.1167/4.12.11

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95. http://dx.doi.org/10.1038/nature04262

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. http://dx.doi.org/10.1038/nature06860