How does episodic memory develop in adolescence?

Imogen R Mechie, Kate Plaisted-Grant, and Lucy G Cheke

Department of Psychology, University of Cambridge, Cambridge CB23EB, United Kingdom

Key areas of the episodic memory (EM) network demonstrate changing structure and volume during adolescence. EM is multifaceted and yet studies of EM thus far have largely examined single components, used different methods and have unsurprisingly yielded inconsistent results. The Treasure Hunt task is a single paradigm that allows parallel investigation of memory content, associative structure, and the impact of different retrieval support. Combining the cognitive and neurobiological accounts, we hypothesized that some elements of EM performance may decline in late adolescence owing to considerable restructuring of the hippocampus at this time. Using the Treasure Hunt task, we examined EM performance in 80 participants aged 10–17 yr. Results demonstrated a cubic trajectory with youngest and oldest participants performing worst. This was emphasized in associative memory, which aligns well with existing literature indicating hippocampal restructuring in later adolescence. It is proposed that memory development may follow a nonlinear path as children approach adulthood, but that future work is required to confirm and extend the trends demonstrated in this study.

Episodic memory (EM) describes the ability to encode, store, and retrieve representations of previously experienced episodes and their temporal-spatial context (Tulving 1972). EM development continues well into the third decade of life (Ruggiero et al. 2016); however, its developmental trajectory after the preschool years remains controversial, with some studies suggesting linear improvements (Ofen et al. 2007) and others no improvement (Picard et al. 2012) or a nonlinear pattern (Tulving 1985; Keresztes et al. 2017). While there has been some debate as to the "defining features" of EM (Cheke and Clayton 2013, 2015) most theorists agree that it is not a unitary ability, instead reflecting the combination of a number of contributing features. Given that many of these studies used different methods for testing EM, and that different tests may emphasize different features (Cheke and Clayton 2013, 2015), it is likely that empirical differences reflect the fact that different features of memory may develop differently during later childhood and adolescence (Picard et al. 2012).

The importance of understanding the developmental trajectory of EM in adolescence is highlighted in the close association between EM and other cognitive processes. EM is thought to support decision-making, particularly in the incorporation of memories into task- and goal-relevant responses (Murty et al. 2016); thus, immaturity of EM may influence the high levels of risk taking observed in adolescence. Adolescence also represents a period of vulnerability to the development of mental illness (Kessler et al. 2007). Evidence that deficits in EM have been linked to a number of mental health disorders such as depression (Goodwin 1997) and anxiety (Airaksinen et al. 2005) raises the possibility that individual differences in memory development during this period may influence this vulnerability. Finally, adolescence is a demanding time academically: During these school years, large quantities of knowledge must be acquired to be successful in exams, which have long-term impacts on individuals' academic and professional future. It is therefore important to understand factors that may contribute to individual differences and challenges in learning and memory during this period.

Memory development in adolescence has attracted considerable research attention in recent years, with the majority of work conducted on developmental trajectories of brain areas within the memory network. EM relies on a distributed network of brain areas, including the medial temporal and superior parietal lobes and the prefrontal cortex (PFC) (Simons and Spiers 2003). Each area within the network, as well as the network itself, shows protracted maturation across adolescence.

Development of the memory network during adolescence

Structural changes in the PFC extend throughout adolescence into adulthood (Spear 2000) and may be nonlinear and multifaceted, with research providing evidence for a peak in gray matter volume at ~11 yr (Giedd et al. 1999) followed by a decrease, while others demonstrate gradual cortical thinning from 7 yr of age (Ducharme et al. 2016; Sowell et al. 2004, 2007). This shift in trajectory of gray matter volume is thought to reflect protracted synaptogenesis, increasing capacity for higher cognitive functions (Huttenlocher and Dabholkar 1997), followed by synaptic pruning of obsolete connections to produce maximally efficient neural pathways (Huttenlocher 1979). According to this account, at peak gray matter volume, large numbers of obsolete connections might feasibly compromise cognitive efficiency. Indeed, there is some evidence that degree of cortical thinning during this period is associated with improved memory recall (Sowell et al. 2001) and this is linked with increased memory-related activity in PFC regions, particularly the dorsolateral PFC (Ofen et al. 2012).

Hippocampal volumes increase throughout childhood (e.g., Brown et al. 2012; Gilmore et al. 2012); however, investigations of its development through adolescence has produced inconsistent findings, with some indicating stable volume (e.g., Koolschijn and Crone 2013), some indicating increases (e.g., Dennison et al. 2013), and others decreases in hippocampal volume during the teenage years (Tamnes et al. 2013). More recent studies suggest a quadratic trajectory of development (e.g., Herting et al. 2018;

Corresponding authors: lgc23@cam.ac.uk, imogen.mechie@nhs.net
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.053264.120.

Tamnes et al. 2018), which may explain some of the earlier inconsistencies. Further inconsistency in this literature may stem from variation in developmental trajectory between different hippocampal subregions, although these studies also show inconsistent findings, likely reflecting variations in sampling (cross-sectional, longitudinal, or accelerated longitudinal) and segmentation techniques. That being said, many of these studies indicate quadratic or cubic development during adolescence in specific subregions (DeMaster et al. 2014; Daugherty et al. 2017; Tamnes et al. 2018). Adding yet another level of complexity, there appear to be changes in the way in which the hippocampus is recruited during memory performance over the period spanning late childhood, adolescence, and early adulthood (DeMaster et al. 2014; Sastre et al. 2016). Finally, the frontal–temporal network, a crucial part of a functioning EM system in adults (Simons and Spiers 2003; Blumenfeld and Ranganath 2007), is also developing during adolescence (e.g., Sherman et al. 2014; Simmonds et al. 2014).

How these neurodevelopmental changes are reflected in memory performance is unclear, as demonstrated by the elaborate patchwork of studies that exist, individually examining aspects of the relationship of frontal or hippocampal structure and functioning in relation to measures of memory. To date, no research has specifically investigated the developmental trajectory of different component of EM within an integrated framework.

Behavioral changes in memory performance across adolescence

There is evidence for a nonlinear developmental trajectory in certain components of EM development. Lee et al. (2014) suggest that performance on associative memory during the middle childhood and adolescent period may be quadratic in nature. They showed that 8- to 9-yr-old children performed significantly more poorly in an item–color associative memory task than 9- to 11- and 13- to 15-yr-old children, but not the intermediate 11- to 13-yr-old. This performance, when controlling for age, was associated with the volume in the right hippocampus (particularly CA3/DG), which also demonstrated a nonlinear developmental pattern during this period, with highest volumes in the 11- to 13-yr-old children. In contrast, tasks that might be considered to preferentially rely on frontal processing for example, assessment of "remembered" as opposed to "familiar" memories show linear improvements between 8 and 24 yr and are associated with functional and structural development of the DLPFC but not with any measure of medial temporal lobe volume (Ofen et al. 2007). Interestingly, these authors note that their results may be "better described in a nonlinear function," but this was not something they assessed.

The heterogeneity of previous data suggests that the trajectory of memory development seen may depend on the nature of memory assessed. Different tasks assessing different components of EM may produce different trajectories, likely reflecting development of different brain areas. In support of this, Keresztes et al. (2017) conducted a number of memory assessments in participants aged 6–14 and 18–27 and found linear improvements in some, such as source memory, which was correlated to "frontal maturity," and quadratic development of others, such as associative recognition that were positively correlated with "hippocampal maturity." Given that 14- to 18-yr-old were not assessed in this study, it is difficult to identify the age of "peak" performance. However, these findings suggest that memory tasks relying more on frontal function may be expected to show linear increases during this period, while those assessing more hippocampal-dependent processes are more likely to show nonlinear development.

The complication of puberty

Adolescence is made unique as a developmental period due to the transformational hormonal, psychological and physical effects of puberty. Pubertal status, independently of age, significantly influences subcortical volumes and is likely to be a key driver in the neural maturation in adolescence (Goddings et al. 2014). In their study, using 711 MRI scans from 275 individuals aged 7–20 yr, Goddings et al. (2014) estimated the volume of subcortical structures. They showed that pubertal development, as assessed by Tanner staging, and chronological age had both independent and interactive influences on volume for the hippocampus, amygdala and putamen in both sexes and the caudate in females. In keeping with this, the neurocognitive data suggests puberty-dependent results in cognition. Indeed, nonlinear development producing cognitive "dips" in later adolescence have been observed in other areas of cognition in a manner that was puberty-dependent. For example facial processing is impaired in older adolescence (McGivern et al. 2002) and puberty rather than age per se is thought to account for these changes (Blakemore 2008). As such in this study, analyses are presented with both the entire cohort and with only peripubescent and postpubescent participants. While this does not explicitly investigate the role of puberty (this is confounded with age in our sample), it allows clarification of developmental patterns when variation due to puberty is reduced.

In summary, areas throughout the EM network demonstrate protracted development throughout the adolescent period. These developments may be nonlinear, with gray matter volumes increasing to a peak and subsequently decreasing in a region-specific manner (Giedd 2004; Gogtay et al. 2004). This nonlinear neural development may be reflected in EM performance, depending on what component processes are challenged by the specific task used. However previously used tasks differ in more than just the type of memory they assess, and evidence for varying trajectories may be related to these "nontarget" differences. It is impossible to extrapolate general trends from such isolated studies, demonstrating the need to investigate the different components of EM within the same integrated framework to allow meaningful conclusions to be drawn.

Assessing the component processes of episodic memory

Different theorists have emphasized different component processes that underpin EM in development (Clayton et al. 2003). Clayton et al. (2003) define three criteria for behavioral demonstrations of EM in children and animals: content, structure, and flexibility. Since EM is spatio–temporal in nature, the content of the memory must include information as to what happened ("What"/item memory), where it happened ("Where"/spatial memory) and when it happened ("When"/temporal memory). However, it is not sufficient for all three of these informational elements to be present—they must be structured in an integrated fashion. Thus, the structure of the memory must be associative. Finally, they argue that the memory must be flexibly accessible to conscious recall, and not a mere response to external stimuli. These latter two features overlap significantly with Shing et al. (2010) two-component framework of EM as consisting of an "associative" and "strategic" component. The following section reviews these three components of content: structure (/association), and flexibility (/strategy), and developmental evidence.

Content: What, Where, and When

The content component of EM concerns remembering information about events (What), locations (Where), and times (When).

In general, these can be translated as item memory, spatial memory and temporal memory.

Studies agree that item memory steadily increases with age up until the eighth year. Beyond this age, some studies show a continued increase (Riggins 2014), others an increase from six to nine and then a plateau (Picard et al. 2012), and others age invariance (Ghetti and Angelini 2008). These differences likely reflect the different stimuli used (e.g., words vs. pictures), and task difficulty. For example, Keresztes et al. (2017) showed a quadratic development of item recognition for faces, while Daugherty et al. (2017) showed no development for word memory over a similar period (six to 27 and eight to 25, respectively). Other studies have demonstrated different developmental trajectories depending on the level of retrieval support (see "Flexibility: Strategic Remembering and Retrieval Support").

Spatial memory appears to be more consistent, with most studies showing linearly increasing ability when sampling between 1 and 20 yr (e.g., Bauer et al. 2012; Ruggiero et al. 2016) with the exception of one study showing evidence of age invariance after 4 yr (Sluzenski et al. 2006).

Temporal memory lags behind item and spatial memory in the early years of life (e.g., Hayne and Imuta 2011; Scarf et al. 2017). However, results on development trajectories after this point have been largely inconsistent. In studies assessing relative recency, some studies have indicated no improvement in memory for item recency between 4 and 18 yr (Brown 1973), while others demonstrated improvement between 5 and 12 yr with age on similar tasks (Mathews and Fozard 1970; Von Wright 1973). Others have argued that different types of temporal memory judgments (relative recency vs. temporal position) develop at different rates, with recency judgements being more easily made by younger children (Friedman 1991, 2013). Memory for temporal location may not be reliable until the age of six (Friedman 1991) but appears to be relatively age invariant beyond this point (Friedman et al. 2010).

All three content features (item, spatial and temporal memory) are thought to rely to various degrees on the medial temporal lobe, but may differ in the extent and nature of hippocampal involvement, with spatial and temporal memory being particularly hippocampal (Burgess et al. 2002; Palombo and Verfaellie 2017). Given this, in the current study we might predict a more nonlinear pattern of development in temporal and spatial memory compared with item memory.

Structure: association

Clayton et al. (2003) emphasize that EM must not merely contain information on item, space, and time, but that this information must be structured as a bound representation. This association of elements is reflected in the "association component" described by Shing et al. (2010).

A large amount of neuroimaging data implicates the hippocampus as being critical for the association of item, spatial and temporal information (e.g., Cheke et al. 2017; Davachi and Wagner 2002; Konkel and Cohen 2009) in order to create a unique episode, which can be differentiated from other similar episodes (Devito and Eichenbaum 2010; Ergorul and Eichenbaum 2004). Given the hypothesis that more hippocampal-dependent elements are more likely to show nonlinear development in the teenage years, what evidence is there of nonlinear development in associative memory in adolescence?

Associative memory can be assessed in many ways: Usually tasks require the association of two features or stimuli, which may either be arbitrarily combined (e.g., two unrelated words presented together), or may form a more coherent unit (e.g., face–name, or a word written in a colored ink). Item–location associative

memory has been shown to improve between 4 and 8 yr (Bauer et al. 2012; Sluzenski et al. 2006) even when accounting for memory for the individual elements. While the evidence seems to consistently report developmental change in associative memory through late childhood and adolescence, some report linear improvements (Daugherty et al. 2017) while others indicate a quadratic developmental trajectory (Keresztes et al. 2017; Lee et al. 2014) and most studies agree that performance in associative memory tasks are linked with maturity of the hippocampal formation.

A number of developmental studies investigating the association between item, spatial, and temporal information have been conducted in recent years. This "What–Where–When" (WWW) memory has been shown to improve with age between 2 and 7 yr (e.g., Cheke and Clayton 2015; Hayne and Imuta 2011; Huttenlocher et al. 2016), but few of these controlled for memory for the individual elements, and none (to our knowledge) extend this investigation beyond the age of 7 yr (although see P Guo, E Carey, K Plaisted-Grant, et al., in prep, for an investigation in middle childhood). Due to the established reliance on hippocampal function, we hypothesize that item–location–time ("What–Where–When") associative memory will demonstrate nonlinear (cubic) development in the 10–17 age range, with the youngest and oldest adolescents being outperformed by those of intermediate age.

Flexibility: strategic remembering and retrieval support

A major source of development in memory from birth to adulthood appears to be in the degree to which retrieval is rigidly dependent on cues from the environment (Gee and Pipe 1995; Usher and Neisser 1993). Memory retrieval can occur as a reflexive response to a familiar stimuli (recognition), in response to external cues that trigger the retrieval of a memory (cued recall) or spontaneously, in response to internally generated cues (free recall). The third and final component of Clayton et al.'s (2003) model of EM is flexibility; the idea that a memory representation must be accessible through self-generated retrieval mechanisms, and available for flexible use in decision-making. Reducing the amount of retrieval support in the form of cues is thought to increase the necessity of episodic recollection, reflecting in evidence that individuals are more likely to report "remembering" items that have been freely recalled as compared with those that have been cued (Tulving 1985; Yonelinas 2002).

Age-related differences during early and middle childhood are more pronounced in situations where less retrieval support is provided (e.g., Paz-Alonso et al. 2009; Cheke and Clayton 2015). Free recall requires more self-initiation and therefore puts higher demand on frontal executive compared with cued recall or recognition (Craik and McDowd 1987; Shing et al. 2010). This self-initiation forms part of what Shing et al. (2010) describe as the "strategic" component of memory, which is concerned with searching, selecting and organizing memory features. This facilitates purposeful encoding strategies, as well as being important for "source monitoring"—that is, remembering the context in which information was learned—both of which demonstrate protracted development (Pressley and Schneider 1997; Keresztes et al. 2017). Like executive functions, with which they overlap, these strategic processes are highly dependent on the prefrontal cortex and in particular the dorsolateral prefrontal cortex (Achim and Lepage 2005; Badre and Wagner 2007; Blumenfeld et al. 2011). Shing et al. (2010) suggest that the framework for strategic memory is established from 10 to 13yr of age but may undergo a "transition period" in which the benefits of strategy use fail to materialize. To our knowledge, there has not been a previous investigation of the impact of retrieval support on

memory performance across the adolescent years. If peak gray matter in the PFC implies that frontal-dependent processes should demonstrate a "dip" very early in the adolescent period (around age 10), we hypothesize that performance advantage afforded by increased retrieval support should gradually—and linearly—decrease during the teenage years,

Assessing multiple elements of EM in a single paradigm: the Treasure Hunt task

From the review above, it is clear that when considering the development of EM, this cannot be seen as a unitary ability, but a multifaceted cognitive process. Studies using different methodologies to assess particular elements of EM demonstrate variance in developmental trajectory (Cheke and Clayton 2013), and comparing between studies, it is difficult to ascertain whether differences seen were due to task demands or other influences. To understand the relative development of different component factors, it is important to investigate these within a single paradigm.

The present study examines the developmental trajectory of EM using a variant of the "Treasure Hunt task" (Cheke et al. 2016), a computer-based task in which participants are presented with scenes and asked to hide objects around the scenes on different days. Following the hiding phase, participants are prompted to remember what they hid (identify previously seen items), where (identify locations used) and when (identify item order) as well as what-where-when combinations (identify the location an item was hidden during a particular time period) with different levels of retrieval support. The Treasure Hunt task enables assessment of individual item, place and time memory ability (content) as well as the ability to integrate these into a single representation (structure/association) within the same paradigm, based on the same encoding phase. In addition, while keeping the encoding constant, the Treasure Hunt task permits manipulation of retrieval support (contrasting recognition and cued-recall tasks) such that flexibility/strategy can also be investigated. Neuroimaging investigation of this task has indicated that the association of elements, rather than individual elements alone elicited activation within the hippocampus and angular gyrus (Cheke et al. 2017). Successful associative memory, but not item memory, was also associated with activity in the dorsolateral prefrontal cortex (DLPFC), with activity in this area during retrieval being associated positively with integrated memory performance, and activity at both encoding and retrieval negatively correlating with binding errors. As such, this task is able to assess multiple elements of memory, as defined from both from a psychological and neuroscientific perspective.

A number of further features make the Treasure Hunt task an attractive tool for measuring EM. Participants are responsible for generating their own associations by hiding items themselves during the encoding phase. This makes encoding closer to "real life" than the arbitrary associations presented in other paradigms and recall is prompted nonverbally using simplistic cues, reducing confounds pertaining to verbal ability. The task has also been validated across a wide age range from middle childhood to old age (Cheke 2016; Silva et al. 2019; P Guo, E Carey, K Plaisted-Grant, et al., in prep.).

In the present study, we investigate multiple components of EM using the Treasure Hunt task in 80 adolescents aged 10–17 yr. Based on previous behavioral data, we predict that some elements of memory will demonstrate linear improvement during this period, while others may demonstrate nonlinear (cubic) development. Given the heterogeneity of previous findings it is difficult to predict the precise pattern of nonlinear development; however, they may broadly tie with the average timing of lobe-specific neural

maturity. Peak gray matter volume (GMV) in the frontal lobe has been suggested to be achieved at ~11 yr (Giedd et al. 1999), whereas peak GMV in the temporal lobe (and the hippocampus) occurs at 17 yr. Following the account that suggests that this cubic trajectory reflects synaptogenesis followed by synaptic pruning of obsolete connections (Peter 1979), we suggest that peak GMV may be reflected in inefficient cognitive performance (McGivern et al. 2002), which may then be followed by improvements as pruning progresses. Based on these timings, we therefore predict that during the 10–17 period, we should see broadly linear increases in performance with age when demands are placed on more frontal processes—for example, the strategic retrieval required with reduced retrieval support (represented in our data by the "support benefit" variable)—while a nonlinear (cubic) pattern may be seen with increased demand on hippocampal functions; that is, spatial, temporal, and associative memory (here represented by the "Where," "When," and "What–Where–When" tasks). Adolescence is a period of change on multiple levels, one of which is pubertal status. In our sample we are unable to independently investigate age and puberty due to the high relatedness of these variables. Instead, we present the main analyses twice, once with the whole sample, and once with only the postpubescent participants, this allows investigation of whether age-related patterns are present when variation due to puberty is reduced, or whether they are reliant on pubertal change per se.

Results

To correct for oversampling of older participants (see Fig. 6), a fractional weighting variable was created based on the expected population proportion for each age group (in years: 12.5%) such that all age groups contributed equally to the analysis. Analyses were then conducted across all participants and again separately, considering only the postpubescent participants. In addition to the regression analyses quoted in the text, all analyses conducted can be seen in Table 1.

Overall EM performance across age

A repeated-measures ANOVA with within-subject factors of Support (two levels: High Support and Low Support) and Task (four levels: What, Where, When, and WWW) against age in months as a covariate reveals a significant main effect of task ($F_{(3,70)} = 19.124$, $P < 0.001$, $\eta^2 = 0.450$), a main effect of support ($F_{(1,72)} = 10.89$, $P = 0.002$, $\eta^2 = 0.131$) and a task × age interaction ($F_{(1,70)} = 6.183$, $P = 0.001$, $\eta^2 = 0.209$). However, there was no main effect of age ($F_{(1,72)} = 9.507$, $P = 0.003$, $\eta^2 = 0.117$), support × age ($F_{(1,72)} = 0.46$, $P = 0.83$, $\eta^2 = 0.001$), support × task ($F_{(2,70)} = 0.975$, $P = 0.409$, $\eta^2 = 0.040$), or support × task × age interaction ($F_{(1,70)} = 1.873$, $P = 0.142$, $\eta^2 = 0.074$). Overall, performance on all four tasks differed significantly from one another, with the "What" task attracting the highest scores, followed by "Where," followed by "WWW" and finally the "When" tasks were found the most difficult (see Fig. 1). Overall "What" scores were significantly higher than all other tasks (all Ps < 0.001), When scores were significantly lower than all other tasks (all Ps < 0.001), and Where scores were significantly higher than WWW scores ($P < 0.001$). All these analyses survived correction for multiple comparisons. Overall, High Support scores were significantly higher than Low Support scores ($P < 0.001$). Finally, High Support tasks attracted significantly higher scores in the When task ($F_{(1,78)} = 8.376$, $P = 0.005$) but not for any of the other individual tasks (WWW: $F_{(1,78)} = 0.041$, $P = 0.840$; What: $F_{(1,78)} = 0.125$, $P = 0.725$; Where: $F_{(1,78)} = 1.322$, $P = 0.254$). The "What" task showed a considerable ceiling effect (38% of cases achieving top score). As such, this task was converted into a binary variable (top score/nonstop scores).

Table 1. Regression analysis r^2, P, and BF values for all regressions conducted on participant performance against age in months

Regression model		Cubic	Linear
Content	What	Binary logistic regression (with binary variable) All participants $\beta(0.008) < 0.001$, $P = 0.971$ Postpuberty only $\beta(0.019) = 0.037$, $P = 0.057$	
	Where	All participants $r^2 = 0.010$, $P = 0.690$ Postpuberty only $r^2 = 0.039$, $P = 0.419$	All participants $r^2 = 0.001$, $P = 0.759$, $BF_{01} = 2.57$ Postpuberty only $r^2 = 0.006$, $P = 0.616$, $BF_{01} = 3.88$
	When	All participants $r^2 = 0.007$, $P = 0.775$ Postpuberty only* $r^2 = 0.132$, $P = 0.046$	All participants $r^2 < 0.001$, $P = 0.977$, $BF_{01} = 7.14$ Postpuberty only $r^2 = 0.019$, $P = 0.366$, $BF_{01} = 1.42$
Structure	Integrated WWW score	All participants* $r^2 = 0.091$, $P = 0.026$ Postpuberty only** $r^2 = 0.180$, $P = 0.014$	All participants* $r^2 = 0.056$, $P = 0.035$, $BF_{10} = 137.46$ Postpuberty only* $r^2 = 0.011$, $P = 0.495$, $BF_{10} = 14.35$
	Nonintegrated/(averaged content) score	All participants $r^2 = 0.004$, $P = 0.855$ Postpuberty only* $r^2 = 0.161$, $P = 0.022$	All participants $r^2 < 0.001$, $P = 0.894$, $BF_{01} = 2.05$ Postpuberty only $r^2 = 0.019$, $P = 0.363$, $BF_{10} = 31.03$
	Structuring difficulty score	All participants** $r^2 = 0.122$, $P = 0.007$ Postpuberty only $r^2 = 0.076$, $P = 0.179$	All participants** $r^2 = 0.083$, $P = 0.010$, $BF_{10} = 100.64$ Postpuberty only $r^2 = 0.002$, $P = 0.780$, $BF_{01} = 2.28$
Flexibility	High Support score	All participants* $r^2 = 0.078$, $P = 0.045$ Postpuberty only** $r^2 = 0.260$, $P = 0.001$	All participants* $r^2 = 0.062$, $P = 0.026$, $BF_{10} = 20.58$ Postpuberty only* $r^2 = 0.074$, $P = 0.067$, $BF_{10} = 2300.89$
	Low Support score	All participants $r^2 = 0.010$, $P = 0.692$ Postpuberty only $r^2 = 0.115$, $P = 0.071$	All participants $r^2 < 0.001$, $P = 0.932$, $BF_{01} = 1.39$ Postpuberty only $r^2 = 0.001$, $P = 0.806$, $BF_{10} = 2.99$
	Support benefit	All participants $r^2 = 0.024$, $P = 0.386$ Postpuberty only $r^2 = 0.029$, $P = 0.524$	All participants $r^2 = 0.024$, $P = 0.171$, $BF_{01} = 10.64$ Postpuberty only $r^2 = 0.012$, $P = 0.475$, $BF_{01} = 7.35$

Significant results ($\alpha = 0.05$) and Bayes factors indicating moderate or higher ($BF > 3$) evidence to support an association with age are signified by a single asterisk. Where results survive adjustment for multiple comparisons (Sidak corrected α: content $\alpha = 0.003125$; structure and flexibility $\alpha = 0.01563$) they are denoted by two asterisks.

Nonparametric analysis revealed no impact of support on this task (Wilcoxen, $W = -0.164$, $P = 0.869$). Repeating the repeated measures ANOVA without the "What" task did not change the pattern of results (with the possible exception of bringing the Support × Task × Age interaction up to a nonsignificant trend $F_{(2,71)} = 2.850$, $P = 0.064$, $\eta^2 = 0.074$).
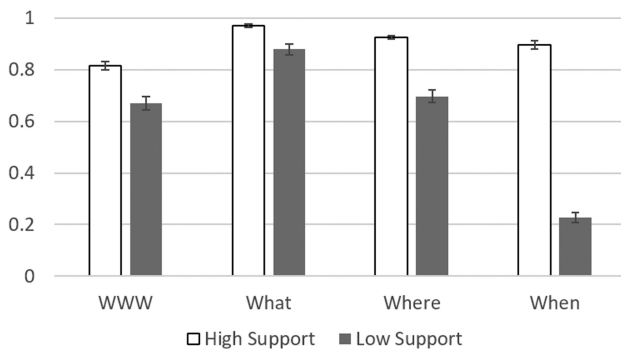

Figure 1. Mean What, Where, When, and WWW scores in the High and Low Support versions of the task.

Content

Regression analysis of the three content elements What (as a binary variable), Where, and When (as continuous variables) against age in months was performed, modeling the data against linear and cubic trajectories (Fig. 2). "What" score did not show a significant binary logistic regression with age (What: all participants $\beta(0.008) < 0.001$, $P = 0.971$), but cubic models could not be assessed. Cubic and linear models were nonsignificant for "Where" and "When" scores suggesting age-invariant performance (Table 1). A JZS Bayesian linear regression with default priors suggested that there was anecdotal ($BF_{01} = 2.57$) and moderate ($BF_{01} = 7.14$) evidence for accepting the null hypothesis of no change with age for Where and When, respectively (Table 1).

Structure/association

Regression analysis on associative memory (Integrated WWW score) demonstrated a significant cubic trajectory (cubic regression: all participants $r^2 = 0.091$, $P = 0.026$). The linear model also demonstrated significance, perhaps capturing the early improvement in performance, and Bayesian analysis suggested this indicated "extreme evidence" (integrated WWW: linear regression: all participants $r^2 = 0.056$, $P = 0.035$, $BF_{10} = 137.46$). However, these analyses did not survive adjustment for multiple comparisons (Sidak $\alpha = 0.01563$).
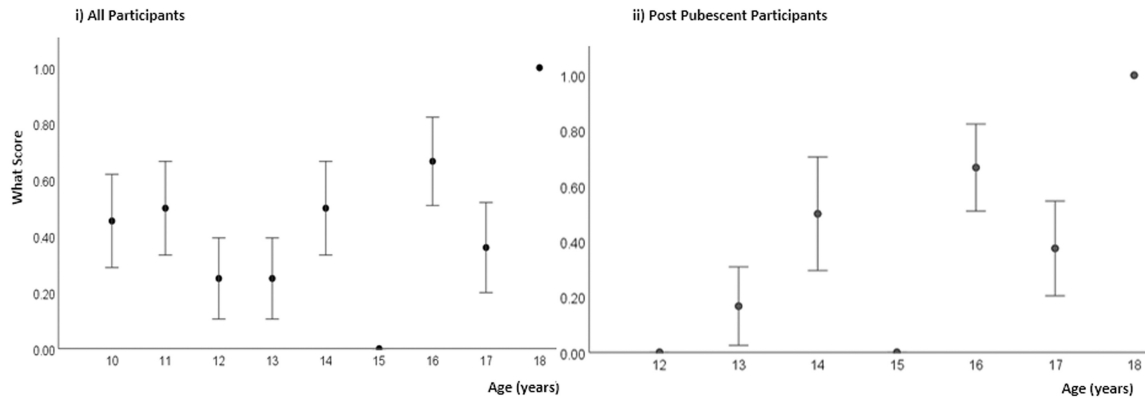
Figure 2. Binary "What" score against age in years with fractional weighting in all participants (*i*) and postpubescent participants (*ii*).

To control for memory for the individual elements, a measure of structuring difficulty was created, by subtracting the integrated score from the averaged content (What, Where, and When) scores. Regression analysis on structure difficulty score showed a significant cubic and linear trajectory across all participants, which survived multiple comparison adjustment (linear regression: all participants: $r^2 = 0.083$, $P = 0.010$; cubic regression: all participants: $r^2 = 0.122$, $P = 0.007$). Bayesian analysis of the linear model suggested extreme evidence to support an association ($BF_{10} = 100.64$). These results suggest the greatest difficulty with associating multiple components in the youngest and oldest participants, and that this was not driven by individual content features (Table 1; Fig. 4).

Flexibility / strategy

The degree to which participants benefited from retrieval support was investigated by calculating an average High Support and an average Low Support score (averaged content + WWW scores in the HS and LS format, respectively). Both linear and cubic regressions of the High Support score were significant when considering all participants (High Support score: linear: all participants: $r^2 = 0.062$, $P = 0.026$; cubic: all participants: $r^2 = 0.078$, $P = 0.045$). Although neither survived adjustment for multiple comparisons (Sidak $\alpha = 0.01563$), the JZS Bayesian analysis suggested that there was strong evidence for the linear model ($BF_{10} = 20.58$). Cubic and linear regressions of the Low Support score were both nonsignificant and Bayesian analysis suggested that there was anecdotal evidence to accept the null hypothesis ($BF_{01} = 1.39$). Support benefit—that is, the degree to which performance was improved in the High Support relative to the Low Support task—was then calculated as the difference between the High and Low Support scores and regression analysis was performed. The support benefit did not appear to be modeled by either linear or cubic models and Bayesian analysis suggested that there was strong evidence to accept the null hypothesis that performance did not change with age ($BF_{01} = 10.64$) (see Table 1; Fig. 5).

Postpubescent data analysis

Given the considerable impact of puberty on brain development, it is important to consider pubertal status. However, the overlap between age and pubertal status in this sample is high, rendering it impossible to compare prepubescent and postpubescent data independently of age. Instead, the same analyses are repeated on only the postpubescent data. This maintains the age range of greatest interest (12–18; $N = 53$) while reducing the confounding influence of pubertal status

Content: postpubescent cohort only

Regression analysis of the three content elements What (as a binary variable), Where, and When (as continuous variables) against age in months was performed, modeling the data against linear and cubic trajectories in the postpuberty cohort. The binary logistic regression of the "What" score was strengthened but remained nonsignificant when considering only postpubescent participants [postpuberty only β(0.019) = 0.037, $P = 0.057$] (Fig. 2).
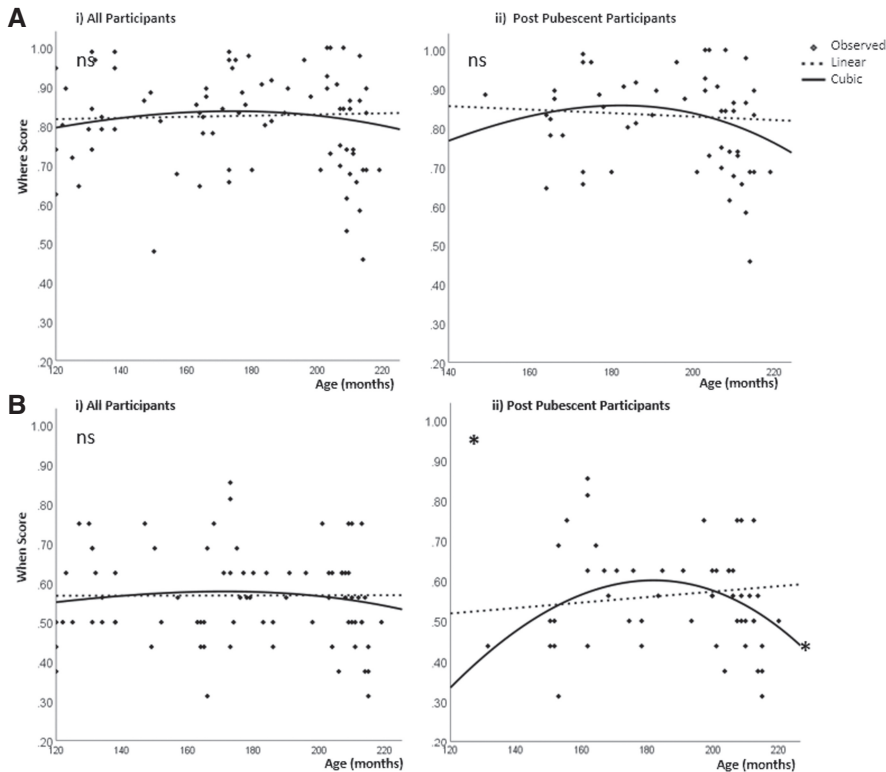
Models for "Where" performance remained nonsignificant in the postpuberty analysis. "When" performance demonstrated a significant cubic model when considering only postpubescent individuals, accounting for 13% of the variance (see Fig. 3) (When: cubic regression: postpuberty only: $r^2 = 0.132$, $P = 0.046$). However, this did not survive the adjustment for multiple comparisons (Sidak $\alpha = 0.03125$). A JZS Bayesian linear regression with default priors suggested there was moderate and anecdotal evidence to accept the null hypothesis for Where ($BF_{01} = 3.88$) and When ($BF_{01} = 1.42$), respectively.

Structure / association: postpubescent cohort only

The significant cubic model observed in the regression analysis of associative memory against age was strengthened when only considering postpubescent participants, surviving the adjustment for multiple comparison (cubic regression: postpuberty $r^2 = 0.180$, $P = 0.014$; Sidak $\alpha = 0.01563$) suggesting a significant increase in early years in association performance and subsequent decrease later in adolescence. The linear model lost significance ($r^2 = 0.011$, $P = 0.495$), but Bayesian analysis suggested there remained strong evidence for the model ($BF_{10} = 14.35$). Regression analysis of the structuring difficulty score lost significance in both the linear and cubic models (linear regression: postpuberty: $r^2 = 0.002$, $P = 0.780$; cubic regression: postpuberty: $r^2 = 0.076$, $P = 0.179$) and Bayesian analysis suggested there was anecdotal evidence to accept the null hypothesis ($BF_{01} = 2.28$) (Table 1; Fig. 4).

Flexibility / strategy: postpubescent cohort only

When considering only the postpubescent cohort, regression analysis of the average High Support score against age in months strengthened the cubic model, accounting for 26% of the variance and withstanding the correction for multiple comparisons (cubic regression: postpuberty: $r^2 = 0.260$, $P = 0.001$; Sidak $\alpha = 0.01563$), whereas the linear model lost significance but remained "extreme evidence" for model according to Bayesian analysis (High Support score: postpuberty: $r^2 = 0.074$, $P = 0.067$, $BF_{10} = 2300.89$). Regression analyses against the Low Support score and support

Figure 3. Where (*A*) and When (*B*) performance as a function of age in months in all participants (*i*) and postpubescent participants (*ii*) modeled against linear and cubic regressions.

benefit remained nonsignificant (see Table 1; Fig. 5). Bayesian analysis suggested there was anecdotal evidence to support a linear model for the low support task ($BF_{10} = 2.99$) and moderate evidence to support the null hypothesis for support benefit ($BF_{01} = 7.35$).

Strategy

Participants were asked to report on what strategies they used in the task. All but two participants (female 120 mo, male 179 mo) reported using strategies to aid memory. ANOVA (IV: strategy; DV: age) performed with the data weighted by age group showed no difference in strategy type used with age ($F_{(2,77)} = 0.304$, $P = 0.583$). There was no association between strategy type and performance (all $Fs < 1$).

Discussion

This study aimed to investigate the developmental trajectory of different elements of EM in a cross-sectional sample of children aged 10–17. We found that while EM appears to show both linear and nonlinear features over this age range depending on the aspect being tested, it was in general better characterized by a cubic model (particularly when there was a high level of retrieval support). The results are broadly consistent with the mixed previous research demonstrating both linear and nonlinear development over the teenage years. Furthermore, these findings tie in well with neurobiological evidence of different developmental trajectories for different neural areas within the EM network. Broadly speaking, the tasks that were predicted to be more hippocampal-dependent, such as temporal and associative (WWW) memory, were more likely to demonstrate (or be better predicted by) a cubic trajectory, with a peak at ~15–16 followed by a considerable

dip in performance at around the age of 17. This timeline reflects some previous behavioral findings (Keresztes et al. 2017) as well as the suggested period of peak gray matter volume of the hippocampus (Giedd et al. 1999).
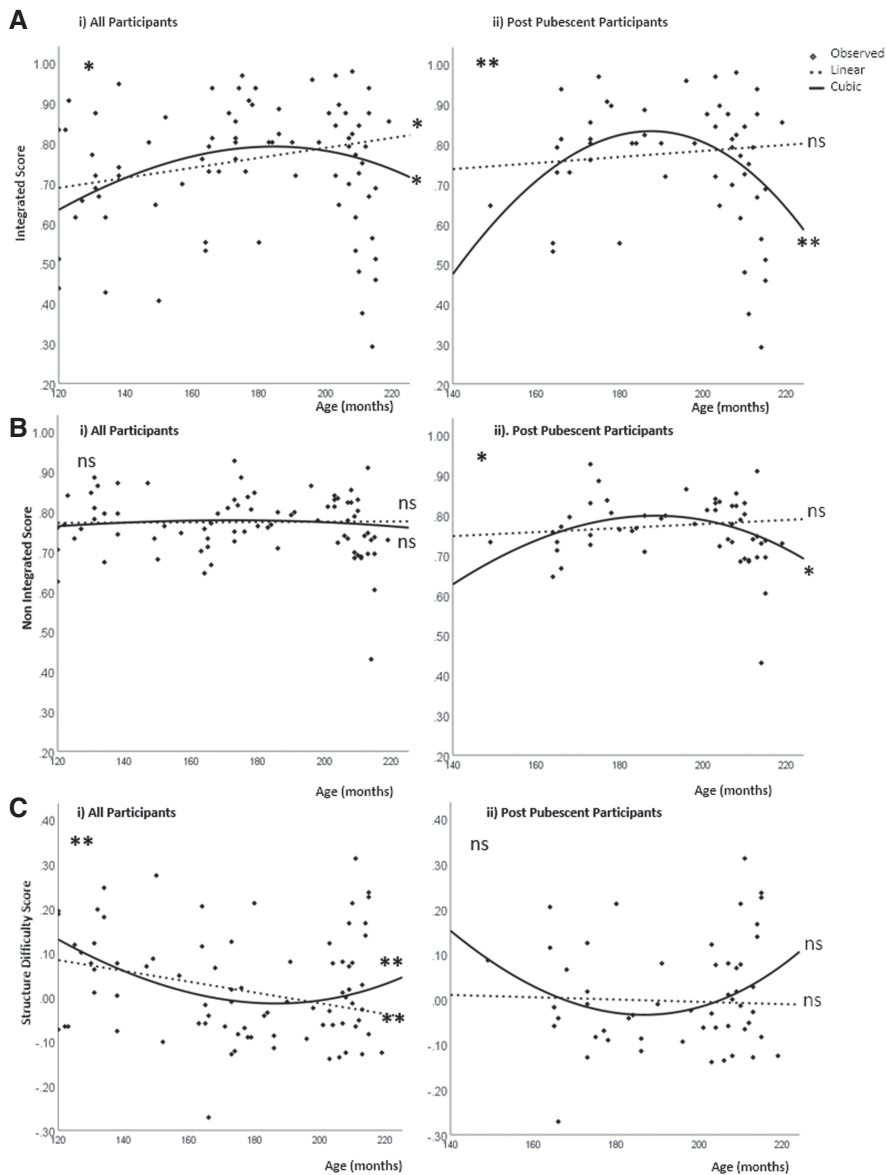
Content

Temporal and Location memory are thought to be more demanding on hippocampal function than item memory (Burgess et al. 2002; Konkel and Cohen 2009; Palombo and Verfaellie 2017). Reflecting our hypothesis that more hippocampal dependent processes would produce more nonlinear trajectories, we observed a significant cubic trajectory in "When" ability considering only postpubescent participants. While the "When" model did not survive correction for multiple comparisons, the distinction in model fit between the linear and cubic models in this test should be noted: The cubic model accounted for ~13% of the variance, in contrast to the linear model, which accounted for <2%. Bayesian analysis provided weak support for the null hypothesis in the linear model, suggesting that the distinction between cubic and linear here was genuine. There is little research currently exploring temporal and spatial memory during adolescence; however, previous studies have suggested no change, or linear improvement throughout development as a whole (Brown 1973; Ruggiero et al. 2016). One explanation for this difference may be the exact developmental period covered by the present study, and the fact that we explicitly investigated nonlinear models, which was not the case in all previous studies. The lack of either linear or cubic change with age in the "Where" task is interesting and unexpected. The Bayesian analysis suggests that there is only weak evidence to accept the null hypothesis here (although this increases to moderate for postpubescent individuals); thus, it is not clear whether this is a "genuine" null result or simply a smaller effect size. Certainly, however, our data do not provide any evidence for a change in spatial memory performance over the adolescent period.

"What" ability showed a ceiling effect with high performance across the 10- to 17-yr range, which makes it difficult to assess trajectory of item memory. This is likely to have arisen for a combination of reasons: First, age invariance in item memory has often been seen after midadolescence in previous studies (Ghetti and Angelini 2008; Picard et al. 2012). Second, a necessary feature of the Treasure Hunt task is that a single encoding event is assessed by multiple retrieval tasks and that the individual content elements are thus the same as those assessed in the association task. This means that in order to keep the difficulty of the association task achievable, the number of item elements must be limited. An unfortunate consequence is that this task often produces a ceiling effect in the "What" task. Such a flaw can be countered by using multiple difficulty levels, as has been shown in studies with different populations (e.g., Cheke et al. 2016; P Guo, E Carey, K Plaisted-Grant, et al., in prep.) and this should be addressed in future work in order to better examine developmental trajectories in item memory in this age group. For the current study, we addressed this by recoding the "What" performance into a binary variable

Figure 4. Associative memory (WWW) (*A*), nonintegrated scores (averaged What Where When scores) (*B*), and the structuring difficulty score (*C*) as a function of age across all participants (*i*) and postpubescent participants (*ii*) modeled with linear and cubic regressions. (*) Significant model fit, (**) significant fit model that survives multiple comparisons.
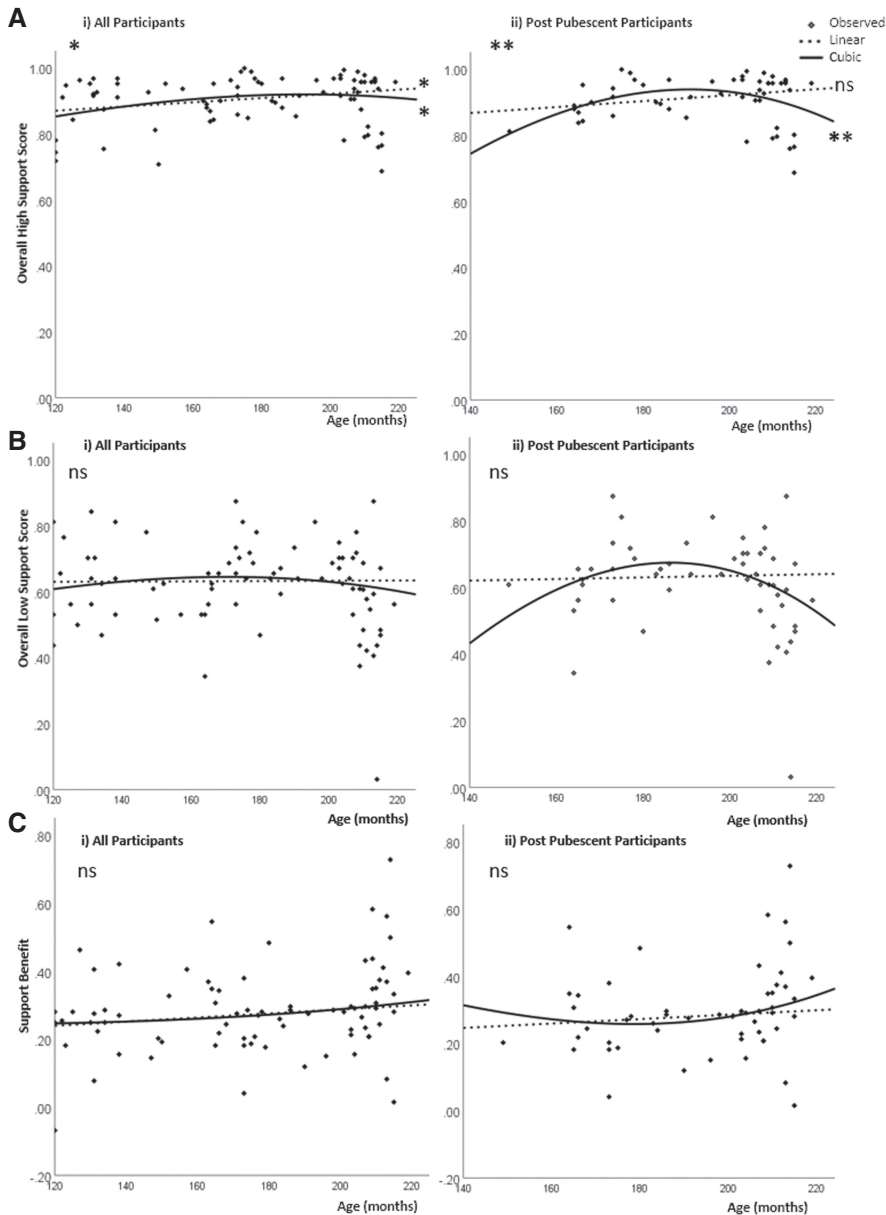
ticipants, with the cubic model strengthened when prepubescent participants were removed. This model survived correction for multiple comparisons and explained 18% of observed variance (compared with the linear model that accounted for only 1%, but was still considered "strong evidence" by the Bayesian analysis). Integrating item memory with temporal and spatial information must rely to some extent on the memory for individual elements (content). To remove this confound and more purely examine association ability, we devised a "structuring difficulty score" by subtracting individuals' average content scores ("nonintegrated score") from the WWW score. There were notable differences in the age-related change in the nonintegrated content score depending on whether prepubescent individuals were included in the analysis. When all participants were considered, the nonintegrated score showed no association with age; however, when only postpubescent individuals were included, the nonintegrated score demonstrated a significant cubic association with age. The linear model lost significance; however, the Bayesian analysis suggested there was still "very strong" evidence for the model. These differences broadly reflect the pattern observed in the three individual content scores, and filter through to the resulting structuring difficulty score: When all participants are considered, structure difficulty shows a highly significant cubic trajectory, with the youngest and oldest participants finding association of elements more difficult than middle adolescent participants. The linear regression is also significant, although it accounts for slightly less of the variance in performance (8% vs. 12% in the cubic trajectory). This suggests that the nonlinear developmental trajectory seen in associative memory may not be due entirely to developmental changes in memory for content. However, when variation due to puberty is removed, this pattern disappears. The role of puberty here is difficult to interpret. It is possible that the difference in the model-fits is due to the inclusion—or not—of prepubescent individuals: It may be that it is the onset of puberty (rather than age per se) that instigates changes in associative memory. It is also possible that it was the inclusion of the younger age groups (10 and 11 yr old) all of whom were prepubescent and therefore not represented in the "postpuberty" group, that influenced this pattern. Future studies de-confounding age and pubertal status are needed to explore this further.

("full marks" and "not full marks"). While this lost some important variance (for example, 15 yr olds scored generally higher than younger children on this task, but none achieved full marks; thus, on the binary variable, it appears that they did poorly), it facilitated analysis demonstrating no significant impact of support, but no improvement with age. It did not, however, allow a cubic model to be explored. Thus, it remains unclear whether item memory is better described by a linear or nonlinear trajectory.

Structure/association

Association of features has been suggested as a key function of the hippocampus (Burgess et al. 2002), which has specifically been shown to be recruited by the integrated WWW element of the Treasure Hunt task (Cheke et al. 2017). Associative (WWW) memory showed significant cubic and linear development across all par-

Flexibility/strategy

Controlling for task, supporting the retrieval significantly improved performances for all ages. Significant cubic and linear trajectories were seen in the high support but not the low support recall formats. When only postpubescent participants were

Figure 5. Overall scores on the High Support tasks (*A*), Low Support tasks (*B*), and support benefit (*C*). Regressions performed on all participants (*i*) and and postpubescent participants (*ii*) modeled against cubic and linear regressions. (*) Significant model fit, (**) significant fit model that survives multiple comparisons.

cence has not, not our knowledge, been previously conducted. It is therefore unclear to what extent our finding of no change in self-generated retrieval across adolescence fits in with existing behavioral work. Given the importance of the DLPFC in retrieval and response monitoring (e.g., McDonough et al. 2013) we might have predicted the degree of support benefit to be related to frontal maturity, which is hypothesized to be improving throughout this period (Giedd et al. 1999; Keresztes et al. 2017). As such it is perhaps surprising to see no change in our sample. One potential explanation is that the same processes underpinning the dip in performance in association ability (i.e., restructuring of the hippocampal formation) undermines or cancels out improvements in self-generated retrieval that might otherwise be seen in older adolescents. Such an account would need to be explored in further research.

Shing et al. (2010) suggest that mnemonic strategy use is first established between the ages of 10 and 13 yr. In our study, all but two participants reported using strategies to aid memory. When weighting our data for age group, the type of strategy used did not significantly differ with age. There was also no relationship with performance. It is likely that having a strategy is not a good enough measure of ability to *use* a strategy effectively, something that was not captured by our measure.

Conclusions and caveats

We believe that this is the first study to investigate the development of the components of EM in the adolescent period from 10 to 17 yr. Due to the nature of this investigation, models were assessed against multiple tasks. This raises the potential of false positives to arise from multiple comparisons, and we have indicated which analyses survive correction for this. However, it was our intention in this study not to focus on any single results but to assess the pattern of findings across tasks and age. On this basis we hypothe-

considered, the cubic model was strengthened and the linear weakened, such that only the postpuberty cubic model survived correction for multiple comparisons, explaining 26% of variance, compared with 7% in the linear model (which nonetheless provides "extreme" evidence to reject the null hypothesis). There was no significant change with age in the difference between the two support tasks (i.e, the extent to which performance is improved in the presence of greater retrieval support), suggesting that effortful retrieval is not something that either improves or declines during this period. Indeed, this was the only area in which the Bayesian analysis indicated strong evidence to accept the null hypothesis of no change over age. A direct investigation of the impact of retrieval support on memory performance in adoles-

sized that tasks considered to be more reliant on hippocampal function would be more likely to demonstrate nonlinear development. We also used a Bayesian linear regression model to give an indication of where the differences in cubic and linear models were because the linear model did not fit the data, and where there was simply a difference in the degree to which the models explained variance. Our results support the hypotheses to some degree: A nonlinear development was seen in some more traditionally hippocampal dependent tasks (temporal and associative, but notably not spatial, memory), which is in keeping with the neurocognitive account of gray matter changes across the memory network, and particularly the hippocampus, during this period. This nonlinearity is particularly notable for temporal memory in

the postpubescent cohort, where the cubic model was significant but the linear model was both nonsignificant and with a low Bayes factor.

Nonetheless, our study suffers from a number of limitations and as such further research will be required before firm conclusions can be drawn. First, like most developmental studies, this investigation was cross-sectional. Longitudinal studies are necessary to fully understand development of cognitive processes over time in a manner that is not confounded with individual differences. Longitudinal investigation would be particularly interesting given the nonlinear development suggested by our data, especially given that we were unable to properly investigate how these changes may interact with pubertal status. Past studies have demonstrated that it is puberty, rather than age, that correlates best with the late adolescent changes observed (Blakemore 2008) and indeed it is noted that the regressions where only postpubescent participants were considered in this study generally strengthened the cubic regression models. However, this study was unable to investigate pubertal status as a variable, and thus it is not possible to know whether it is exclusion of younger participants, or puberty itself that influences the difference between models. Future studies should also consider more sophisticated means of assessing pubertal stage than the binary presence or absence of secondary hair growth, which creates a false "threshold" of puberty in place of the gradual change seen in reality. Such a measurement (alongside explicit recruitment strategies) would allow for pubertal status to be modeled as a covariate against age. While 80 is a reasonable sample size, the distribution of participant ages raises the possibility of skew in the results obtained: Cubic patterns may have been seen due to a greater variability in the older age groups due to a larger sample size rather than genuinely lower performance. Our analysis accounted for this by weighting the data such that each age group contributed equally; however, replication with an increased and more evenly distributed sample is warranted, and this too would be addressed in longitudinal design. A further issue is that we were able to assess the strength of evidence for the linear models using a Bayesian analysis, but this was not straightforward for nonlinear or binary logistic analyses. This means that we were not able to directly compare the strength of evidence for linear and nonlinear models. Finally, while we have linked the current findings to both behavioral and neuroscientific literature, conclusions about the neural underpinnings of the developmental patterns seen in our data cannot be confidently drawn without concurrent investigation of neural development in the same participants. Future investigations should combine our novel behavioral paradigm with structural and functional scanning techniques, to comprehensively investigate how neural development influences the development of difference aspects of EM across adolescence.

In summary, we have demonstrated that different elements of EM demonstrate different developmental trajectories across adolescence. Broadly speaking, we predicted that elements that are thought to be more hippocampal dependent, such as spatial, temporal, and associative memory, would be likely to demonstrate nonlinear development, reflecting restructuring of the hippocampal formation during this period. In line with our hypotheses, temporal and associative memory demonstrated significant cubic trajectories, with reduced performance in older participants; however, spatial memory did not. Item memory, which is thought to be less hippocampal dependent, did not demonstrate significant age-related change, but due to this needing to be recoded as a binary variable, it was not possible to assess a cubic model for this. High support forms of the memory tasks were more likely to demonstrate significant age-related change (with the cubic models being stronger). However, the extent to which participants benefit from retrieval support did not change during this period. That the timing of the cognitive "dip" in performance in older adoles-

cence aligns with the average age of peak GMV in the hippocampus is of note particularly because neural inefficiency associated with peak GMV has often been linked more with changes in neural activity and processing speed rather than in task performance (e.g., DeMaster et al. 2014; Sastre et al. 2016). Further investigation using longitudinal neuroimaging is required to ascertain how these behavioral patterns are related to developmental changes in neural structure and engagement.

Our study suggests that previous discrepancies in behavioral results regarding the trajectory of memory development may have arisen due to measuring different components of EM. EM relies on a range of interacting component processes, as well as a widely distributed network of brain areas. It is therefore unsurprising that different types of challenge would produce different developmental findings, especially during times of considerable neural reorganization such as adolescence. If borne out through future studies, evidence of reduced EM ability in late adolescence may be of considerable significance. EM is being increasingly recognized as an important factor in decision-making (Murty et al. 2016) and mental health disorders (Goodwin 1997), both of which are core areas of research in adolescence, where risky decisions and vulnerable mental health are key challenges to wellbeing. Furthermore, late adolescence is a time at which individuals are under considerable academic pressure, taking exams that will have significant impact on their future professional opportunities. For all of these reasons, understanding the nature of memory development throughout adolescence is crucial if we are to support healthy and successful development in the transition to adulthood.

Materials and Methods

Participants

Eighty participants (female $n = 34$, male $n = 46$) aged 10–17 yr (male: M = 173.13 mo, SD = 30.00 mo; female: M = 182.50 mo, SD = 32.92 mo) (see Fig. 6). were recruited from a range of UK state and independent schools by means of flyers, emails and posters. Their date of birth was recorded and age on testing day calculated to the nearest month. Written consent was obtained from each participant and a parent/guardian before partaking in the study. Where participants had to travel to the testing location, they were remunerated to reflect the costs incurred. This study received ethical approval from the Cambridge Psychology research Ethics Committee.

Pubertal status

The development of axillary hair growth occurs with the onset of andrenarche. It can be characterized using Wolfsdorf staging, a noninvasive method of assessing pubertal status in adolescents. Self-reported presence of axillary hair was used to characterize participants as either stage 1 (prepubertal) or stage 2+ (peripubertal and postpubertal).

The Treasure Hunt task

The Treasure Hunt task, devised by Cheke et al. (2016), is a What–Where–When style memory task that permits simultaneous assessment of Content (individual What, Where, and When), Structure (What–Where–When binding) and Accessibility (self-generation ability).

In the treasure hunt task, each participant undergoes a brief training session where they are presented with a complex virtual scene on a computer screen and then asked to "hide" an everyday item somewhere in the scene. They hide two versions of each item, one on each of two "days" presented consecutively and then asked to remember where they hid each item, and indicate this by placing each item in the same location they previously placed it. Feedback is given based on whether they placed each item in the correct location for each "day." Following the training, four
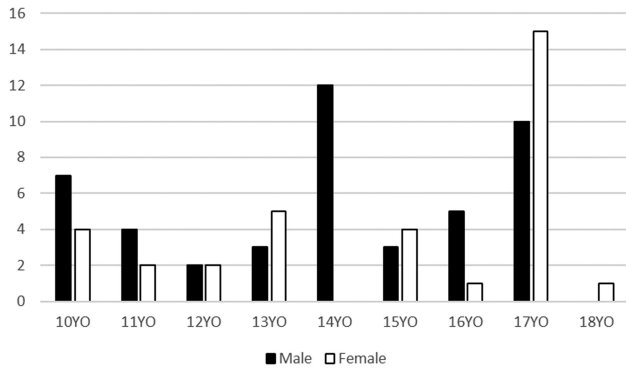
Figure 6. Number of participants in each age group

sessions were administered to each participant counterbalanced between participants to prevent order effects

The sessions differ in their retrieval support: two are "High Support" (HS1 and HS2) and two "Low Support" tasks (LS1 and LS2). Two versions of each session were presented (e.g. LS1 vs. LS2) these took the same format but differed in the scenes and items presented. All participants completed LS1 and LS2; however, there were three files corrupted in result extraction process, two from LS1 and 1 from LS2 making a total of 157 out of 160 results. During the initial stages of the data collection process, one of the HS2 sessions malfunctioned, and thus 27 participants carried out only HS1, with 53 participants carrying out both HS1 and HS2. As there was no significant difference within participants between their score on LS1 versus LS2 and HS1 versus HS2, these were averaged. Where only one data set was present, this score was taken as their "average score."

Each session had an encoding and retrieval phase. During the encoding phase, participants were asked to hide two items (e.g., a chocolate bar and a can of drink) around two complex scenes (e.g. a common room and a yard). Each item was hidden twice, across two immediately consecutive time-periods (clearly labeled "day 1" or "day 2"). Participants moved items using the arrow keys, pressing "enter" to hide the item in a place of their choosing within the scene, having full autonomy over their hiding behavior. Each participant performed eight hiding events per session, reflecting eight unique item–location–day combinations (e.g., item 1–scene 1–day 1, item 2–scene 1–day 1, item 1–scene 1–day 2, etc.). All sessions (LS1, LS2, HS1, and HS2) had the same encoding format but scenes and items changed between sessions (see Fig. 7). For each session, at a fixed time interval after the encoding period (∼5 min), the participant was asked to recall their hiding behaviors using either a High or Low support retrieval method.

High support

The high support session was a series of recognition tasks wherein participants were presented with binary choices. For "What" memory, they were presented with a series of items, half of which were previously hidden and half of which were novel distractors, and asked "Did you hide this?" to which they indi-

cated yes/no using arrow keys. For "Where" memory, they were presented with a cross in a location on a scene that was either a location in which they previously hid an item, or a random location, and asked to indicate yes/no to the question "Did you hide something here?" For "When" memory, they were presented with two previously hidden items and asked "Which did you hide first?" Finally, for WWW memory participants were presented with ready-made item–location–time associations (i.e., an item placed in a location, with the day clearly indicated) and asked to indicate yes/no to the question "Is that where you hid that item on that day?" (Fig. 7B). In the high support format, What, Where, When, and WWW scores were calculated by the proportion of correct acceptances or rejections. With the exception of the WWW task, these tasks were identical to those used in Cheke et al. (2016).

Low support

The low support session was a series of cued recall tasks wherein participants were required to indicate the correct answers from an array of available responses. Here, "What" memory was assessed by presenting the participant with a range of items and asking them to select which ones they hid by moving a square curser. "Where" memory was assessed by asking participants to place a cross in all the locations where they hid any item (regardless of what the item was or when) in each scene. "When" memory was assessed by presenting icons representing each scene labeled "1" or "2." Participants were asked for each item to move it to the icon representing the scene and serial position in which they previously experienced it (for example moving the first item hidden in scene 1 to the "scene 1" icon with a "1" on it). For WWW memory, participants were asked to "rehide" items in the correct location in
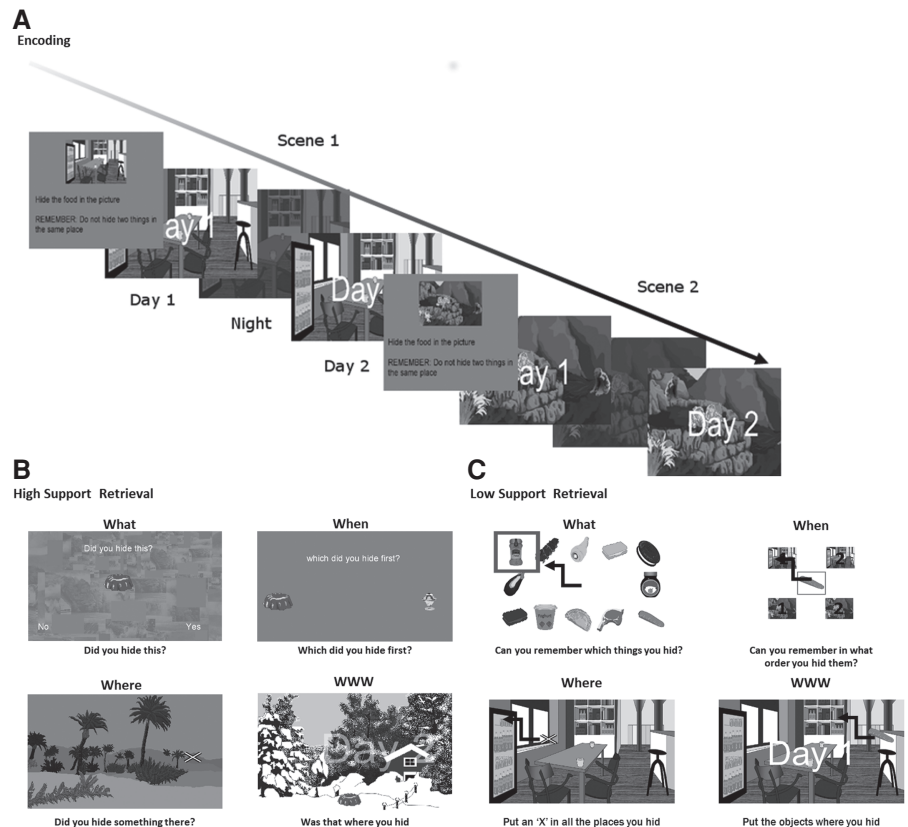
Figure 7. (A) Encoding phase. The participant is asked to hide two items around two scenes over two separate encoding periods labeled "day 1" and "day 2." (B) HS retrieval phase for What, Where, When, and WWW. (C) LS retrieval phase for What, Where, When, and WWW.

the scene on the correct day. For WWW and "Where" memory, scores are calculated by the proportion of spatially matching responses between encoding and recall. For "What" and "When" memory, scores were calculated by the proportion of correct items or icons selected (Fig. 7C).

Measuring episodic memory: content, structure, and flexibility

Content

A single score for each individual element ("What," "Where," and "When") was calculated by averaging the individual scores on that task on the high and low support sessions (e.g. "What" = HS What + LS What/2).

Structure / association

An *Integrated score* was measured by averaging the high and low support "WWW" task scores. To investigate association ability while controlling for memory for the individual elements, a non-integrated score was created, which averaged across the content scores (What + Where + When/3) from which the integrated score was subtracted to create a structure difficulty score. For this score, higher numbers indicate greater difficulty. As such, a score of 0 implies that a participant's ability to integrate What, Where, and When information is as good as their ability to remember individual What, Where, and When information, and there is no "cost" to integration. A negative score implies that integrating features is easier than remember individual features alone. A positive score implies that combining features is more challenging than memory for individual features.

Flexibility / strategy

In this study, flexibility is measured in two ways. First, it is defined by the degree to which participants benefited from increased retrieval support. To investigate this, What, Where, When, and WWW scores were averaged in the high support and low support format to calculate a single "High Support" and "Low Support" score. Support benefit—that is, the degree to which performance was improved in the High Support relative to the Low Support task—was then calculated as the difference between these two scores. Thus, a higher support benefit indicates that an individual may rely more heavily on external cues and has less "flexible" or "strategic" retrieval ability. Additionally, after completing the tasks, participants were asked "Did you have a strategy for remembering where and when you hid items?" and "Can you explain it to me?" Their answers were coded as being "spatial" if they were hidden based on screen position (e.g. "I always hid items on day 1 on the left and day 2 on the right") or "salience" if hiding places were chosen based on screen content (e.g. "I hid the items in obvious places like the bottle on top of the table").

Analysis

For each element of EM—content, structure, and flexibility—we investigated how performance differed between participants as a function of age using regression analysis, ANOVA and paired t-tests conducted on IBM SPSS with significance reported at $\alpha = 0.05$. Where necessary, Sidak correction for multiple comparisons was used. To assess strength of evidence of the linear models, JZS Bayesian linear regressions with default priors was conducted. A Bayes factor of three or more was considered at least moderate evidence, either for (BF_{10}) or against (BF_{01}) an effect. As many psychological and neural changes occur at puberty, we subsequently performed the same analysis removing Wolfsdorf stage 1 participants to consider only pubescent/postpubescent participants (Wolfsdorf sStage 2+).

References

Achim AM, Lepage M. 2005. Dorsolateral prefrontal cortex involvement in memory post-retrieval monitoring revealed in both item and associative recognition tests. *Neuroimage* **24:** 1113–1121. doi:10.1016/j.neuroimage.2004.10.036

Airaksinen E, Larsson M, Forsell Y. 2005. Neuropsychological functions in anxiety disorders in population-based samples: evidence of episodic memory dysfunction. *J Psychiatr Res* **39:** 207–214. doi:10.1016/j.jpsychires.2004.06.001

Badre D, Wagner AD. 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* **45:** 2883–2901. doi:10.1016/j.neuropsychologia.2007.06.015

Bauer PJ, Doydum AO, Pathman T, Larkina M, Güler OE, Burch M. 2012. It's all about location, location, location: children's memory for the 'where' of personally experienced events. *J Exp Child Psychol* **113:** 510–522. doi:10.1016/j.jecp.2012.06.007

Blakemore SJ. 2008. The social brain in adolescence. *Nat Rev Neurosci* **9:** 267. doi:10.1038/nrn2353

Blumenfeld RS, Ranganath C. 2007. Prefrontal cortex and long-term memory encoding: an integrative review of findings from neuropsychology and neuroimaging. *Neuroscientist* **13:** 280–291. doi:10.1177/1073858407299290

Blumenfeld RS, Parks CM, Yonelinas AP, Ranganath C. 2011. Putting the pieces together: the role of dorsolateral prefrontal cortex in relational memory encoding. *J Cogn Neurosci* **23:** 257–265. doi:10.1162/jocn.2010.21459

Brown AL. 1973. Judgments of recency for long sequences of pictures: the absence of a developmental trend. *J Exp Child Psychol* **15:** 473–480. doi:10.1016/0022-0965(73)90097-0

Brown TT, Kuperman JM, Chung Y, Erhart M, McCabe C, Hagler DJ, Venkatraman VK, Akshoomoff N, Amaral DG, Bloss CS, et al. 2012. Neuroanatomical assessment of biological maturity. *Curr Biol* **22:** 1693–1698. doi:10.1016/j.cub.2012.07.002

Burgess N, Maguire EA, O'Keefe J. 2002. The human hippocampus and spatial and episodic memory. *Neuron* **35:** 625–641. doi:10.1016/S0896-6273(02)00830-9

Cheke LG. 2016. What–where–when memory and encoding strategies in healthy aging. *Learn Mem* **23:** 121–126. doi:10.1101/lm.040840.115

Cheke LG, Clayton NS. 2013. Do different tests of episodic memory produce consistent results in human adults? *Learn Mem* **20:** 491–498. doi:10.1101/lm.030502.113

Cheke LG, Clayton NS. 2015. The six blind men and the elephant: are episodic memory tasks tests of different things or different tests of the same thing? *J Exp Child Psychol* **137:** 164–171. doi:10.1016/j.jecp.2015.03.006

Cheke LG, Simons JS, Clayton NS. 2016. Higher body mass index is associated with episodic memory deficits in young adults. *Q J Exp Psychol* **69:** 2305–2316. doi:10.1080/17470218.2015.1099163

Cheke LG, Bonnici HM, Clayton NS, Simons JS. 2017. Obesity and insulin resistance are associated with reduced activity in core memory regions of the brain. *Neuropsychologia* **96:** 137–149. doi:10.1016/j.neuropsychologia.2017.01.013

Clayton NS, Bussey TJ, Dickinson A. 2003. Can animals recall the past and plan for the future? *Nat Rev Neurosci* **4:** 685–691. doi:10.1038/nrn1180

Craik FIM, McDowd JM. 1987. Age differences in recall and recognition. *J Exp Psychol Learn Mem Cogn* **13:** 474–479. doi:10.1037/0278-7393.13.3.474

Daugherty AM, Flinn R, Ofen N. 2017. Hippocampal CA3-dentate gyrus volume uniquely linked to improvement in associative memory from childhood to adulthood. *Neuroimage* **153:** 75–85. doi:10.1016/j.neuroimage.2017.03.047

Davachi L, Wagner AD. 2002. Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *J Neurophysiol* **88:** 982–990. doi:10.1152/jn.2002.88.2.982

DeMaster D, Pathman T, Lee JK, Ghetti S. 2014. Structural development of the hippocampus and episodic memory: developmental differences along the anterior/posterior axis. *Cerebral Cortex* **24:** 3036–3045. doi:10.1093/cercor/bht160

Dennison M, Whittle S, Yücel M, Vijayakumar N, Kline A, Simmons J, Allen NB. 2013. Mapping subcortical brain maturation during adolescence: evidence of hemisphere- and sex-specific longitudinal changes. *Dev Sci* **16:** 772–791. doi:10.1111/desc.12057

Devito LM, Eichenbaum H. 2010. Distinct contributions of the hippocampus and medial prefrontal cortex to the 'what–where–when' components of episodic-like memory in mice. *Behav Brain Res* **215**: 318–325. doi:10.1016/j.bbr.2009.09.014

Ducharme S, Albaugh MD, Nguyen T-V, Hudziak JJ, Mateos-Pérez JM, Labbe A, Evans AC, Karama S. 2016. Trajectories of cortical thickness maturation in normal brain development: the importance of quality control procedures. *Neuroimage* **125**: 267–279. doi:10.1016/j.neuroimage.2015.10.010

Ergorul C, Eichenbaum H. 2004. The hippocampus and memory for 'what,' 'where,' and 'when'. *Learn Mem* **11**: 397–405. doi:10.1101/lm.73304

Friedman WJ. 1991. The development of children's memory for the time of past events. *Child Dev* **62**: 139–155. doi:10.2307/1130710

Friedman WJ. 2013. The development of memory for the times of past events. In *The Wiley handbook on the development of children's memory* (ed. Bauer PJ, Fivush R) pp. 394–407. John Wiley & Sons Ltd., Chichester.

Friedman D, de Chastelaine M, Nessler D, Malcolm B. 2010. Changes in familiarity and recollection across the lifespan: an ERP perspective. *Brain Res* **1310**: 124–141. doi:10.1016/j.brainres.2009.11.016

Gee S, Pipe ME. 1995. Helping children to remember: the influence of object cues on children's accounts of a real event. *Dev Psych* **31**: 746–758. doi:10.1037/0012-1649.31.5.746

Ghetti S, Angelini L. 2008. The development of recollection and familiarity in childhood and adolescence: evidence from the dual-process signal detection model. *Child Dev* **79**: 339–358. doi:10.1111/j.1467-8624.2007.01129.x

Giedd JN. 2004. Structural magnetic resonance imaging of the adolescent brain. *Ann N Y Acad Sci* **1021**: 77–85. doi:10.1196/annals.1308.009

Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL. 1999. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci* **2**: 861–863. doi:10.1038/13158

Gilmore JH, Shi F, Woolson SL, Knickmeyer RC, Short SJ, Lin W, Zhu H, Hamer RM, Styner M, Shen D. 2012. Longitudinal development of cortical and subcortical gray matter from birth to 2 years. *Cereb Cortex* **22**: 2478–2485. doi:10.1093/cercor/bhr327

Goddings AL, Mills KL, Clasen LS, Giedd JN, Viner RM, Blakemore SJ. 2014. The influence of puberty on subcortical brain development. *Neuroimage* **88**: 242–251. doi:10.1016/j.neuroimage.2013.09.073

Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, Nugent TF III, Herman DH, Clasen LS, Toga AW, et al. 2004. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci* **101**: 8174–8179. doi:10.1073/pnas.0402680101

Goodwin GM. 1997. Neuropsychological and neuroimaging evidence for the involvement of the frontal lobes in depression. *J Psychopharmacol* **11**: 115–122. doi:10.1177/026988119701100204

Hayne H, Imuta K. 2011. Episodic memory in 3- and 4-year-old children. *Dev Psychobiol* **53**: 317–322. doi:10.1002/dev.20527

Herting MM, Johnson C, Mills KL, Vijayakumar N, Dennison M, Liu C, Goddings AL, Dahl RE, Sowell ER, Whittle S, et al. 2018. Development of subcortical volumes across adolescence in males and females: a multisample study of longitudinal changes. *Neuroimage* **172**: 194–205. doi:10.1016/j.neuroimage.2018.01.020

Huttenlocher PR. 1979. Synaptic density in human frontal cortex: developmental changes and effects of aging. *Brain Res* **163**: 195–205. doi:10.1016/0006-8993(79)90349-4

Huttenlocher PR, Dabholkar AS. 1997. Regional differences in synaptogenesis in human cerebral cortex. *J Comp Neurol* **387**: 167–178. doi:10.1002/(SICI)1096-9861(19971020)387:2<167::AID-CNE1>3.0.CO;2-Z

Huttenlocher PR, Dabholkar AS, Giedd JN, Davachi L, Wagner AD, Konkel A, Pressley M, Schneider W. 2016. Do different tests of episodic memory produce consistent results in human adults? *Cereb Cortex* **4**: 637–648.

Keresztes A, Bender AR, Bodammer NC, Lindenberger U, Shing YL, Werkle-Bergner M. 2017. Hippocampal maturity promotes memory distinctiveness in childhood and adolescence. *Proc Natl Acad Sci* **114**: 9212–9217. doi:10.1073/pnas.1710654114

Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. 2007. Age of onset of mental disorders: a review of recent literature. *Curr Opin Psychiatry* **20**: 359–364. doi:10.1097/YCO.0b013e32816ebc8c

Konkel A, Cohen NJ. 2009. Relational memory and the hippocampus: representations and methods. *Front Neurosci* **3**: 166–174. doi:10.3389/neuro.01.023.2009

Koolschijn PCMP, Crone EA. 2013. Sex differences and structural brain maturation from childhood to early adulthood. *Dev Cogn Neurosci* **5**: 106–118. doi:10.1016/j.dcn.2013.02.003

Lee JK, Ekstrom AD, Ghetti S. 2014. Volume of hippocampal subfields and episodic memory in childhood and adolescence. *Neuroimage* **94**: 162–171. doi:10.1016/j.neuroimage.2014.03.019

Mathews ME, Fozard JL. 1970. Age differences in judgments of recency for short sequences of pictures. *Dev Psychol* **3**: 208–217. doi:10.1037/h0029582

McDonough IM, Wong JT, Gallo DA. 2013. Age-related differences in prefrontal cortex activity during retrieval monitoring: testing the compensation and dysfunction accounts. *Cereb Cortex* **23**: 1049–1060. doi:10.1093/cercor/bhs064

McGivern RF, Andersen J, Byrd D, Mutter KL, Reilly J. 2002. Cognitive efficiency on a match to sample task decreases at the onset of puberty in children. *Brain Cogn* **50**: 73–89. doi:10.1016/S0278-2626(02)00012-X

Murty VP, Calabro F, Luna B. 2016. The role of experience in adolescent cognitive development: integration of executive, memory, and mesolimbic systems. *Neurosci Biobehav Rev* **70**: 46–58. doi:10.1016/j.neubiorev.2016.07.034

Ofen N, Kao Y-C, Sokol-Hessner P, Kim H, Whitfield-Gabrieli S, Gabrieli JDE. 2007. Development of the declarative memory system in the human brain. *Nat Neurosci* **10**: 1198–1205. doi:10.1038/nn1950

Ofen N, Chai XJ, Schuil KDI, Whitfield-Gabrieli S, Gabrieli JDE. 2012. The development of brain systems associated with successful memory retrieval of scenes. *J Neurosci* **32**: 10012–10020. doi:10.1523/JNEUROSCI.1082-11.2012

Palombo DJ, Verfaellie M. 2017. Hippocampal contributions to memory for time: evidence from neuropsychological studies. *Curr Opin Behav Sci* **17**: 107–113. doi:10.1016/j.cobeha.2017.07.015

Paz-Alonso P, Ghetti S, Matlen BJ, Anderson MC, Bunge SA. 2009. Memory suppression is an active process that improves over childhood. *Front Hum Neurosci* **3**: 24. doi:10.3389/neuro.09.024.2009

Peter RH. 1979. Synaptic density in human frontal cortex: developmental changes and effects of aging. *Brain Res* **163**: 195–205.

Picard L, Cousin S, Guillery-Girard B, Eustache F, Piolino P. 2012. How do the different components of episodic memory develop? Role of executive functions and short-term feature-binding abilities. *Child Dev* **83**: 1037–1050. doi:10.1111/j.1467-8624.2012.01736.x

Pressley M, Schneider W. 1997. *Introduction to memory development during childhood and adolescence*. Erlbaum, Mahwah, NJ.

Riggins T. 2014. Longitudinal investigation of source memory reveals different developmental trajectories for item memory and binding. *Dev Psychol* **50**: 449–459. doi:10.1037/a0033622

Ruggiero G, D'Errico O, Iachini T. 2016. Development of egocentric and allocentric spatial representations from childhood to elderly age. *Psychol Res* **80**: 259–272. doi:10.1007/s00426-015-0658-9

Sastre M, Wendelken C, Lee JK, Bunge SA, Ghetti S, Ghetti S. 2016. Age- and performance-related differences in hippocampal contributions to episodic retrieval. *Dev Cogn Neurosci* **19**: 42–50. doi:10.1016/j.dcn.2016.01.003

Scarf D, Boden H, Labuschagne LG, Gross J, Hayne H. 2017. 'What' and 'where' was when? Memory for the temporal order of episodic events in children. *Dev Psychobiol* **59**: 1039–1045. doi:10.1002/dev.21553

Sherman LE, Rudie JD, Pfeifer JH, Masten CL, McNealy K, Dapretto M. 2014. Development of the default mode and central executive networks across early adolescence: a longitudinal study. *Dev Cogn Neurosci* **10**: 148–159. doi:10.1016/j.dcn.2014.08.002

Shing YL, Werkle-Bergner M, Brehmer Y, Müller V, Li S-C, Lindenberger U. 2010. Episodic memory across the lifespan: the contributions of associative and strategic components. *Neurosci Biobehav Rev* **34**: 1080–1091. doi:10.1016/j.neubiorev.2009.11.002

Silva JSCD, Barbosa FF, Fonsêca ÉKGD, Albuquerque FDS, Cheke LG, Fernández-Calvo B. 2019. Load effect on what-where-when memory in younger and older adults. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* **27**: 841–853. doi:10.1080/13825585.2019.1700207

Simmonds DJ, Hallquist MN, Asato M, Luna B. 2014. Developmental stages and sex differences of white matter and behavioral development through adolescence: a longitudinal diffusion tensor imaging (DTI) study. *Neuroimage* **92**: 356–368. doi:10.1016/j.neuroimage.2013.12.044

Simons JS, Spiers HJ. 2003. Prefrontal and medial temporal lobe interactions in long-term memory. *Nat Rev Neurosci* **4**: 637–648. doi:10.1038/nrn1178

Sluzenski J, Newcombe NS, Kovacs SL. 2006. Binding, relational memory, and recall of naturalistic events: a developmental perspective. *J Exp Psychol Learn Mem Cogn* **32**: 89–100. doi:10.1037/0278-7393.32.1.89

Sowell ER, Delis D, Stiles J, Jernigan TL. 2001. Improved memory functioning and frontal lobe maturation between childhood and adolescence: a structural MRI study. *J Int Neuropsychol Soc* **7**: 312–322. doi:10.1017/S135561770173305X

Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, Toga AW. 2004. Longitudinal mapping of cortical thickness and brain growth in normal children. *J Neurosci* **24**: 8223–8231. doi:10.1523/JNEUROSCI.1798-04.2004

Sowell ER, Thompson PM, Tessner KD, Toga AW. 2007. Mapping continued brain growth and gray matter density reduction in dorsal frontal cortex: inverse relationships during postadolescent brain maturation. *J Neurosci* **21:** 8819–8829. doi:10.1523/JNEUROSCI .21-22-08819.2001

Spear LP. 2000. The adolescent brain and age-related behavioral manifestations. *Neurosci Biobehav Rev* **24:** 417–463. doi:10.1016/ S0149-7634(00)00014-2

Tamnes CK, Walhovd KB, Dale AM, Østby Y, Grydeland H, Richardson G, Westlye LT, Roddey JC, Hagler DJ, Due-Tønnessen P, et al. 2013. Brain development and aging: overlapping and unique patterns of change. *Neuroimage* **68:** 63–74. doi:10.1016/j.neuroimage.2012.11 .039

Tamnes CK, Bos MGN, van de Kamp FC, Peters S, Crone EA. 2018. Longitudinal development of hippocampal subregions from childhood to adulthood. *Dev Cogn Neurosci* **30:** 212–222. doi:10.1016/j.dcn.2018 .03.009

Tulving E. 1972. Episodic and semantic memory. In *Organisation of memory*, 1st ed. (ed. Tulving E, Donaldson W), pp. 381–403. Academic Press, London.

Tulving E. 1985. Memory and consciousness. *Can Psychol* **26:** 1–12. doi:10 .1037/h0080017

Usher JA, Neisser U. 1993. Childhood amnesia and the beginnings of memory for four early life events. *J Exp Psychol Gen* **122:** 155–165. doi:10 .1037/0096-3445.122.2.155

Von Wright JM. 1973. Judgment of relative recency: developmental trends. *J Psychol* **84:** 3–12. doi:10.1080/00223980.1973.9915625

Yonelinas AP. 2002. The nature of recollection and familiarity: a review of 30 years of research. *J Mem Lang* **46:** 441–517. doi:10.1006/jmla.2002 .2864