Development of human hippocampal subfield microstructure related to associative inference

Vinci-Booher, S.* [1, 2], Schlichting, M. [3, 4], Preston, A. [4], & Pestilli, F. [4]

[1] Indiana University, Bloomington, Indiana, United States
[2] Vanderbilt University, Nashville, Tennessee, United States
[3] University of Toronto, Toronto, Canada
[4] University of Texas at Austin, Austin, Texas, United States
* corresponding author

Key words

diffusion, hippocampus, inference, memory, subfields

Abstract

The hippocampus is a complex brain structure composed of subfields that each have distinct cellular organizations. While the volume of hippocampal subfields displays age-related changes that have been associated with inference and memory functions, the degree to which the cellular organization within each subfield is related to these functions throughout development is not well understood. We employed an explicit model testing approach to characterize the development of tissue microstructure and its relationship to performance on two inference tasks, one that required memory (memory-based inference) and one that could be completed based on perceptually available information (perception-based inference). We found that each subfield had a unique developmental trajectory in terms of its cellular organization. While the SUB displayed a linear trajectory, DG, CA1, and CA2/3 subfields displayed non-linear trajectories that interacted with sex in CA2/3. We found that the DG was related to memory-based inference performance and that the SUB, a subfield outside of the hippocampus proper, was related to perception-based inference; neither relationship interacted with age. Results suggest that cellular organization within hippocampal subfields undergo distinct developmental trajectories that may underlie changes in inference and memory performance throughout development.

Development of human hippocampal subfield microstructure related to associative inference

The hippocampus is a complex brain structure that is routinely found to be associated with learning and memory throughout the lifespan (Hirshhorn et al. 2012; Schlichting, Zeithamova, and Preston 2014; Zeithamova, Dominick, and Preston 2012; Barron, Dolan, and Behrens 2013; Scoville and Milner 1957; Lister and Barnes 2009; Scahill et al. 2003; Venkatesh et al. 2020). The structure of the hippocampus contains distinct areas, called subfields, that are distinguishable from one another based on the properties and organization of the cells within each subfield (Manns and Eichenbaum 2006; Lavenex, Banta Lavenex, and Amaral 2007), anatomical differences that are perhaps related to subfield-specific gene expression (Datson et al. 2004; Thompson et al. 2008). Most anatomical studies relating hippocampal subfields to memory throughout the human lifespan have indexed the cellular properties of subfields by measuring the volume of each subfield (Schlichting et al. 2017; Riggins et al. 2015; Daugherty, Flinn, and Ofen 2017; Lee, Ekstrom, and Ghetti 2014; Pereira et al. 2014). Diffusion imaging provides more specific measures of cellular properties than volume that can index changes in the orientation and density of neurons within a particular volume (Callow, Canada, and Riggins 2020), yet the developmental trajectories of diffusion-derived metrics of cellular organization within hippocampal subfields and their relationships to memory throughout the lifespan are unknown. Here, we used diffusion imaging to characterize the developmental trajectories of microstructural changes within each subfield and the relationship of subfield microstructure to memory throughout development.

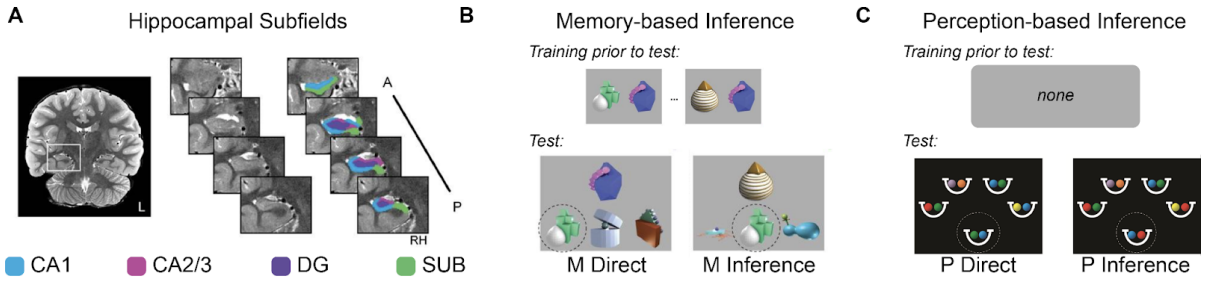

Figure 1. Hippocampal subfields and behavioral tasks. A. Hippocampal subfields. The hippocampus can be segmented into four distinct subfields: CA1 (blue), CA2/3 (pink), DG (purple), and SUB (green). **B. Memory-based inference.** In the memory-based Direct and Inference tasks, participants learned pairs of objects and were later required to select the object from an array of three objects that they learned with a target object (M Direct) and to select the object from an array of three objects that should go with the target object based on a shared feature (M Inference). **C. Perception-based inference.** In the perception-based Direct and Inference tasks, participants were asked to indicate whether or not the target set of balls (emphasized here with a circle) was a valid pair based on the other pairs of balls present on the screen (P Direct) and based on a shared feature (P Inference). Images A and B were adapted from (Schlichting et al. 2017); image C was adapted from (Wendelken and Bunge 2010). CA1 = *cornu ammonis* field 1, CA2/3 = *cornu ammonis* fields 2 and 3, DG = dentate gyrus, SUB = subiculum, A = anterior; P = posterior.

The volume of the hippocampus and its subfields changes throughout the lifespan (Schlichting et al. 2017; Riggins et al. 2015; Daugherty, Flinn, and Ofen 2017; Lee, Ekstrom, and Ghetti 2014; Pereira et al. 2014). In childhood, hippocampal volume displays clear age-related

increases (Hamer et al. 2008) with sex-specific trends (Giedd et al. 1996; Ostby et al. 2009), however there is less consensus on the trajectory of hippocampal volume in adolescence and young adulthood with some studies reporting increases, others reporting decreases, and still others reporting no change, see (Lee, Johnson, and Ghetti 2017) for review. Measuring age-related changes in the volume of the entire hippocampus may mask underlying trends in its anatomically distinct subfields (**Figure 1A**). Indeed, studies that investigated the development of hippocampal subfields, as opposed to the entire hippocampus, revealed that the volume of each subfield undergoes distinct developmental trajectories (Tamnes et al. 2018; Lee, Johnson, and Ghetti 2017; Daugherty, Flinn, and Ofen 2017; Lee, Ekstrom, and Ghetti 2014). While CA1 volume displays a non-linear inverted U-shaped trajectory, CA2/3 and DG volumes display linear downward trajectories, and SUB volume displays a more complex trajectory with an early peak and a late valley. Thus, the volume of hippocampal subfields have distinct developmental trajectories, suggesting that the cellular changes occurring within each subfield might also undergo distinct developmental trajectories.

An emerging body of work is using diffusion imaging to provide information about the development of cellular organization within the hippocampus and its relationship to memory. Diffusion-derived metrics of the hippocampus have developmental trajectories and behavioral correlates that are independent from their volumes (Callow, Canada, and Riggins 2020; Lee, Ekstrom, and Ghetti 2014; Pereira et al. 2014; Wolf et al. 2015; Langnes et al. 2020). For example, mean diffusivity (MD), an index of cellular density (Assaf and Pasternak 2008; P. J. Basser, Mattiello, and LeBihan 1994; Pierpaoli and Basser 1996; Le Bihan 2014), within the hippocampus decreased from 4 years to 13 years of age (Callow, Canada, and Riggins 2020; Mah, Geeraert, and Lebel 2017) and was related to performance on a source memory task (Callow, Canada, and Riggins 2020), even after controlling for hippocampal volume. In a lifespan study of hippocampal *subregions* (i.e., anterior and posterior hippocampus), results indicated different developmental trajectories for MD and volume, with sub-region-specific trends in MD and age-related relationships between sub-region MD and recall memory (Langnes et al. 2020). Investigations of diffusion metrics in hippocampal *subfields* in healthy, non-geriatric populations are still rare to non-existent. Consequently, the developmental trajectories of cellular organization within subfields and their relationships to memory development are nearly entirely based on volume measurements.

Evidence from mouse models suggests that measuring fractional anisotropy (FA), a diffusion-derived metric, may provide important information about the cellular organization within the hippocampus. First, histological work revealed that hippocampal subfields have characteristic fiber orientations. The predominant orientation of diffusion is dorsal-ventral in the DG, rostral-caudal in the CA3 subfield, and either dorsal-ventral or rostral-caudal depending on location within the CA1 subfield (Sierra et al. 2015). Second, experience-induced changes in fiber orientation within subfields can be captured using diffusion-derived estimates of FA. Epileptic injury resulted in an increase in the number of fibers with similar orientations (i.e., a move towards more cellular organization) in the DG and CA3bc subfields, which was consistent with an observed increase in FA (Sierra et al. 2015; Salo et al. 2017). Third, experience-dependent changes in cellular fiber orientations (indexed by FA) can occur within

hippocampal subfields with little to no changes in neuronal density (indexed by MD), suggesting that FA and MD are indexing different tissue properties. While epileptic injury resulted in an increase in FA within DG and CA3bc subfields, there was no observable change in MD in either subfield (Sierra et al. 2015). Furthermore, human MRI studies have demonstrated the feasibility and utility of using FA to measure cellular changes within human hippocampal subfields. FA can be measured reliably in the human hippocampus using diffusion imaging (Müller et al. 2006) and provides information about cellular organization that is independent of MD (Treit et al. 2018; Shereen et al. 2011). FA of the entire hippocampus changes with age (Carlesimo et al. 2010; Venkatesh et al. 2020) and changes with spatial navigation learning in adult humans (Iaria et al. 2008), suggesting that, as in mouse models, changes in the FA of hippocampal subfields may be related to behavior and experience.

Nearly all studies investigating memory integration in the hippocampus have used associative inference paradigms. Associative inference paradigms require participants to learn two independent associations among three objects, such as A goes with B and B goes with C, and they must later infer that A goes with C (Schlichting et al. 2017). Associative inference, therefore, requires successful association learning (i.e., memory of the two pairs) and also successful inference (i.e., reasoning about the unshared features based on a shared feature). According to the memory integration proposal, hippocampal encoding mechanisms accomplish association learning in the context of prior memories, such that successful association learning encodes a neural representation for the learned pairs that would allow for successful inference, whether or not successful inference actually occurs. The ability to make successful inferences may, therefore, rely on additional neural processing that underlies reasoning about learned associations, an ability that can be assessed using “perception-based” inference tasks (Brainerd and Reyna 1992). Perception-based inference tasks require participants to make the same inference as associative inference paradigms, that A goes with C, but the inference does not require long-term memory of the two pairs because AB and BC are both visible during inference (Wendelken and Bunge 2010).

The developmental trajectory of associative inference is more similar to the trajectory of memory for learned associations than perception-based inference, suggesting that neural processing for inference when memory is required may be performed differently than inference when memory is not required. Associative inference ability and memory for learned associations both emerge in the first two to three years of life (Cattell 1987; Halford 1998) and continue to mature throughout childhood and perhaps adolescence (Richland, Morrison, and Holyoak 2006; Sternberg and Rifkin 1979; Vodegel Matzen and Others 1994; Schlichting et al. 2017). Perception-based inference, on the other hand, appears to develop much earlier. By the age of 4 years old, children are capable of perception-based reasoning with novel items (Brainerd and Reyna 1992), suggesting that the protracted development of associative inference is not likely related to the ability to make inferences, per se. Rather, the protracted development of associative inference is likely related to the development of association learning or the development of efficient hippocampal memory integration.

The present work characterizes the developmental trajectories of the tissue microstructure of hippocampal subfields and provides information concerning the relationship of subfield tissue properties and learning and memory. We used diffusion imaging to assess the developmental trajectories of hippocampal subfields based on their tissue microstructure, including CA1, CA2/3, DG, and SUB. More specifically, we employed an explicit model testing paradigm to model the relationship between tissue microstructure and age, from the ages of 6 - 30 years. We then performed separate regressions to predict performance on two inference tasks using subfield microstructure, one that required memory (i.e., associative inference; **Figure 1B**) and one did not require memory (i.e., perception-based inference, **Figure 1C**). For both regressions, we included performance on a control task. For associative inference, the control task was performance on memory for learned pair associations. For perception-based inference, the control task was essentially a matching task for pair associations. This approach allowed us to disentangle relationships between subfields and memory, inference, and associative inference throughout development. If a relationship exists between the tissue properties of hippocampal subfields and performance on the associative inference task but not the perception-based inference task, this would provide support for a relationship between the cellular organization within that subfield and memory integration processing, because memory-integration is not necessary to perform perception-based reasoning (although see (Munnelly and Dymond 2014; Wimmer and Shohamy 2012)). The subfield may facilitate associative inference by supporting efficient encoding and/or retrieval of learned associations as opposed to supporting a separate process of inference.

Materials and methods

Participants

A total of 90 participants were recruited for this study, originally reported in (Schlichting et al. 2017): 31 children (ages 6-11 years), 25 adolescents (ages 12-17 years), and 34 adults (ages 18-30). Only participants that met inclusion several exclusion criteria were used for analysis: (1) absence of psychiatric conditions, i.e., Child Behavioral Check List in children (CBCL; (Achenbach and Edelbrock 1991)) and Symptom Checklist 90-Revised in adults (SCL-90-R, (Derogatis 1977)) in the normal range and (2) average intelligence, i.e., Weschler Abbreviated Scale of Intelligence, Second Edition (Wechsler 2018), full Scale IQ composite score above 2 standard deviations below the mean of a normative sample, resulting in the removal of 1 child, 1 adolescent, and 10 adults, and a remaining sample size of 78 participants (30 children, 24 adolescents, 24 adults). All procedures were conducted under the approval of the Institutional Review Board at the University of Texas at Austin. Parental consent was obtained for participants under 18 years of age. Participants received compensation for their participation.

Hippocampal segmentations for a total of 62 of these participants were acquired from https://osf.io/m87fd/; only participants with available hippocampal subfield segmentations were included in the brain and brain-behavior analyses, leaving 62 participants (22 children, 21 adolescents, 19 adults). Of the 62 participants, an additional 6 participants (1 child, 2 adolescents, 3 adults) were removed due to an absence of one of the required image types (t1, t2, and dwi) or substandard data quality, resulting in 58 participants (21 children, 19 adolescents, 16 adults).

Finally, only participants that completed the MRI session and both behavioral assessments were included in the brain-behavioral analyses, resulting in the removal of an additional 2 participants and a final sample size of 56 participants: 21 children (M = 8.76 years, SD = 1.73 years, Range = [6, 11], 8F, 13M), 19 adolescents (M = 13.79 years, SD = 1.65 years, Range = [12, 17], 9F, 10M), and 16 adults (M = 23.50 years, SD = 3.31 years, Range = [18, 28], 9F, 7M).

Outlier removal for tissue microstructural analyses

Diffusion data from individual participants were removed if their framewise displacement (FD) for the diffusion scan was greater than 1 mm or if their data were determined to be of low signal-to-noise ratio (SNR). This resulted in the removal of two participants (**Supplementary Figure 1**). There were no participants that were removed based on low SNR (**Supplementary Figure 2**). We also identified and removed outliers for each region-of-interest (ROI) according to statistical criteria by applying the box plot rule to the raw residuals and robust weights (Frigge, Hoaglin, and Iglewicz 1989) for each of the models tested (see **Table 2**); these data points were removed selectively for particular ROIs. The final participant counts are provided in **Table 2** and individual participant data are displayed for each model in **Figure 2**.

Outlier removal for behavioral analyses

Behavioral data from individual tasks performed by a participant were removed if they did not understand the task or were below chance on accuracy based on a binomial test for accuracy. The final participant counts are provided for each behavioral measure in **Figure 3** (Memory-based inference: M Matching and M Inference) and **Figure 4** (Perception-based inference: P Matching and P Inference).

Outlier removal for brain-behavior analyses

We identified and removed multivariate outliers for each model tested (see T**able 3**). Multivariate outliers were identified using the box plot rule applied to the raw residuals as well as the robust weights (Frigge, Hoaglin, and Iglewicz 1989). The final participant counts are provided for each model in **Table 3**.

Procedure

Participants completed two visits. During the first visit, participants completed assessments that screened for psychiatric conditions, i.e., Child Behavioral Check List in children (CBCL; (Achenbach and Edelbrock 1991)) and Symptom Checklist 90-Revised in adults (SCL-90-R,(Derogatis 1977)), and potential learning disorders, i.e., Weschler Abbreviated Scale of Intelligence, Second Edition (WASI-II, (Wechsler 2018)), and were familiarized with the MRI environment in a short session in a mock MRI scanner. Participants also completed a battery of tasks designed to assess associative inference and related abilities, always completed in this order: associative inference (Preston et al. 2004; Schlichting et al. 2017; Schlichting and Preston 2015), Iowa Gambling (Bechara et al. 1994), statistical learning (Schapiro et al. 2016; Fiser and Aslin 2001; Schlichting et al. 2017), relational reasoning (Crone et al. 2009), and relational integration (Wendelken and Bunge 2010). The current work focuses on the associative inference and relational integration task because both tasks assess inference and are well matched in terms of task demands. The associative inference task assesses inference when memory is required (i.e., associative inference) whereas the relational integration task assesses inference when memory is not required (i.e., perception-based inference).

Behavioral assessments

Memory-based inference: Matching and Inference

Participants completed an inference task that required memory. This task has been described in detail in (Schlichting et al. 2017) using stimuli that have also been described in (Schlichting and Preston 2015). The task is a modified version of the task described in (Preston et al. 2004) and the stimuli were a modified version of the stimuli described in (Hsu, Schlichting, and Thompson-Schill 2014).

In brief, participants were presented with pairs of objects during a learning phase in pseudo-random order followed by a testing phase (**Figure 1B**). The 30 pairs that were split into two types of pairs: AB pairs and BC pairs. The B object was the same object for each pair type. During the learning phase, participants were encouraged to encode the object pairs by creating stories, either visual or verbal, but were told neither that the B object linked A and C objects nor

that they would be asked to make an inference judgment. During the testing phase, participants were asked to complete a three-alternative forced-choice (3-AFC) task, self-paced, in which they were presented with one of the objects that they had learned and asked to identify the object that was next to it during learning, the Matching task. The cue object was presented at the top and the three choices were presented below. Distractor objects were objects that they had learned but that did not go with the cue object. Trial order was pseudo-randomized and no feedback was provided. The learning and test phases were repeated 4 times each. After the 4 blocks of learning/testing, participants completed the Inference task. The Inference task was only completed once. During Inference, participants were alerted that A and C objects were linked by B items and asked to complete another self-paced 3-AFC test in which they were presented with the C object at the top and asked to select the A object from 3 objects displayed below it. All tasks were all completed using a computer and responses were provided with key presses. We will refer to the memory-based tasks as M Matching and M Inference.

Perception-based inference: Matching and Inference

Participants completed an inference task that did not require memory. This task is described in detail in (Wendelken and Bunge 2010). Note that the full set of tasks includes General Direct, General Inference, Specific Direct, and Specific Inference; however, we use only the General Direct (Matching) and General Inference (Inference) in the current work because it is so well-matched to the memory-based Matching and Inference tasks (described above; **Figure 1B**). In brief, participants were presented with 4 pairs of colored balls, e.g., red-blue, green-purple, yellow-white, blue-orange, and were presented simultaneously with a fifth set of colored balls, the target pair, e.g., red-orange, and asked to make a timed two-alternative forced-choice (2-AFC) judgment about the validity of the fifth set of colored balls given the associations provided in the 4 pairs (**Figure 1C**). For the Matching task, participants were able to determine the validity of the target pair by finding if that pair existed in any of the 4 pairs of colored balls. For the Inference task, participants were required to make an inference using two of the 4 pairs of colored balls presented on the screen. If, for example, a red-blue pair and a blue-orange pair exist among the 4 pairs of colored balls, then the target pair of red-orange would be considered a valid pair. We will refer to the perception-based tasks as P Matching and P Inference.

MRI Data Acquisition and Analyses

MRI data acquisition

A T1-weighted 3D MPRAGE and at least two oblique coronal T2-weighted images were acquired on a 3.0T Siemens Skyra MRI. The T1-weighted image included 256 x256 x192 1 mm^3 voxels. The T2-weighted image was an oblique coronal image acquired perpendicular to the main axis of the hippocampal complex: TR = 13150 ms, TE 82 ms, 512 x 60 x 512 matrix, 0.4 x 0.4 mm in-plane resolution, 1.5 mm thru-plane resolution, 60 slices, no gap. If visual inspection of one of the two T2-weighted images indicated motion artifacts, a third T2-weighted image was collected. Two T2-weighted images were coregistered using ANTS (Avants et al. 2011) and averaged to produce one T2-weighted image for each participant. Diffusion and T2-weighted images were aligned to the T1-weighted image.

Anatomical (T1w) processing

Anatomical images were aligned to the AC-PC plane with an affine transformation using HCP preprocessing pipeline (Glasser et al. 2013) as implemented in the HCP AC-PC Alignment App on brainlife.io (Hayashi, McPherson, and Caron 2018). Images from child subjects were aligned using a pediatric atlas created from 4.5-8.5 year old children (Fonov et al. 2011); images from adult subjects were aligned to the standard MNI152 adult template (Glasser et al. 2013). AC-PC aligned images were then segmented using the Freesurfer 6.0 (Fischl 2012) as implemented in the Freesurfer App on brainlife.io (Hayashi, Kitchell, and Pestilli 2017) to generate the cortical volume maps with labeled cortical regions according to the Destrieux 2009 atlas (Destrieux et al. 2010).

Diffusion (dMRI) processing

All diffusion preprocessing steps were performed using the recommended mrtrix3 preprocessing steps (Ades-Aron et al. 2018) as implemented in the Mrtrix3 Preprocess App on brainlife.io (McPherson 2018). PCA denoising and Gibbs deringing procedures were performed first. We then corrected for susceptibility- and eddy current-induced distortions as well as inter-volume subject motion by interfacing with FSL tools via the *dwipreproc* script. The eddy current-corrected volumes were then corrected for bias field and rician noise. Finally, the preprocessed dMRI data and gradients were aligned to each participant's ACPC-aligned anatomical image using boundary-based registration (BBR) in FSL (Greve and Fischl 2009).

Identification of hippocampal subfields

Hippocampal subfields were identified by hand on each participant's mean coronal image as described in (Schlichting et al. 2017, 2019). Subfields examined included: *cornu ammonis* fields 1 (CA1) and ⅔ (CA2/3), dentate gyrus (DG), and subiculum (SUB).

Microstructural measurements

We estimated fractional anisotropy (FA) and mean diffusivity (MD) based on the tensor model (P. J. Basser, Mattiello, and LeBihan 1994) using the FSL DTIFIT App on brainlife.io (brainlife.app.137). Values for each microstructural measurement were extracted from hippocampal subfields using the Extract diffusion metrics inside ROIs (DTI) App on brainlife.io (brainlife.app.283).

Data and software availability

Data, description of analyses, and web-links to the open source code and open cloud services used in this study are listed in **Table 1** and can be viewed in their entirety here: https://doi.org/10.25663/brainlife.pub.33. The raw data and hippocampal segmentations were transferred to the referenced brainlife.io project from the Open Science Framework (OSF): https://osf.io/m87fd/. Customized code for statistical analyses is available here: https://github.com/svincibo/devti_devHPCsubfields.

Table 1. Analysis steps

Application	Github repository	Open Service DOI
SNR Calculation	https://github.com/davhunt/app-snr_in_cc	https://doi.org/10.25663/brainlife.app.120
dMRI Preprocessing	https://github.com/brain-life/app-mrtrix3-preproc	https://doi.org/10.25663/bl.app.68
FSL DTIFIT	https://github.com/brainlife/app-fslDTIFIT	https://doi.org/10.25663/brainlife.app.137
Extract diffusion metrics	https://github.com/brainlife/app-extract-diffusion-metrics-rois	https://doi.org/10.25663/brainlife.app.283

Statistical analyses

Modeling to determine developmental trajectory

To determine the developmental trajectory of hippocampal microstructure, we assessed the relationship between hippocampal microstructure, age, and sex by testing seven different models. Models were tested for the tissue microstructure of each of the subfields (i.e., CA1, CA23, DG, SUB), for a total of 28 models. For each model, the dependent variable was always the mean fractional anisotropy (FA) of voxels included in the region of interest (ROI), collapsed across hemispheres. Models included: (1) ROI ~ age, (2) ROI ~ age^2, (3) ROI ~ sex, (4) ROI ~ sex + age, (5) ROI ~ sex + age^2, (6) ROI ~ sex * age, (7) ROI ~ sex * age^2. Continuous variables were mean centered. We identified and removed multivariate outliers for each model separately. Multivariate outliers were identified using the box plot rule applied to the raw residuals as well as the robust weights (Frigge, Hoaglin, and Iglewicz 1989). We compared the seven models for each ROI using the Akaike Information Criterion corrected for sample size (*AICc*) to account for the different numbers of predictors in each model (Hurvich and Tsai 1991). The model with the highest *AICc* was selected as the winning model, provided that the model itself was significant, $p < 0.05$. We included sex in our models because subfield volume has been shown to depend on sex (Tamnes et al. 2018).

Supplementary analyses included testing a series of modifications to the response variable, including: ROIs that included the hippocampal regions instead of subfields (i.e., head, body, tail), testing the intersection of subfields and subregions ROIs (i.e., CA1_head, CA23_head, DG_head, SUB_head, CA1_body, CA23_body, DG_body, SUB_body), and testing an ROI that included the entire hippocampus as well as testing a microstructural model other than FA (i.e., mean diffusivity (MD)). We note that MD is commonly used in subcortical gray matter regions, such as the hippocampus (Peter J. Basser and Jones 2002).

Analysis of behavioral data

Each behavioral measure was analyzed separately using Two-way Repeated Measures ANOVAs with two factors: Age and Task. Age always had three levels: children, adolescents, adults. Task always had two levels. Task had two levels for the memory-based tasks, Matching and Inference; Task had two levels for the perception-based tasks, Matching and Inference. The dependent variable was either performance (i.e., accuracy) or reaction time. Pearson correlations were used to evaluate the presence of a speed-accuracy trade-off for each task (**Supplementary Figure 3**). One-sample *t*-tests were performed to confirm that performance was above chance for each task within each age group. Chance was 33% for the memory-based tasks and 50% for perception-based tasks.

Determining the relationships among inference, memory, and the microstructure of hippocampal subfields

To determine the relationship between the tissue microstructure of hippocampal subfields and memory performance as a function of age, we used multiple linear regression to assess the degree to which task performance on each task could predict subfield microstructure. In order to compare memory-based and non-memory based inference, we fit separate models for each subfield ROI that included a memory-based (Matching and Inference) and a non-memory-based task (Matching and Inference). Given the degree of correlation between tasks within the memory-based tasks and the non-memory-based tasks (**Supplementary Table 1**), we conducted two separate regressions for each ROI, one that included the matching tasks (memory-based Matching and perception-based Matching) and a second that included the inference tasks (memory-based Inference and perception-based Inference). We also included the model parameters found in the prior analysis that best explained the relationship between that subfield's microstructure, age, and sex. Finally, we included IQ as a covariate of no interest to account for general intelligence, as in prior work (Schlichting et al. 2017). While accuracy and reaction time for the final repetition on the memory-based Matching task was evaluated in the behavioral analyses estimating learning, we found that the high performance on the final repetition of the memory-based Matching task left the measurement with a low level of variability and, therefore, elected to use the accuracy on the memory-based Matching task averaged across all four repetitions, as in prior work (Schlichting et al. 2017). Multivariate outliers were identified and removed for each model separately. Multivariate outliers were identified using the box plot rule applied to the raw residuals as well as the robust weights (Frigge, Hoaglin, and Iglewicz 1989). Continuous variables were mean centered. Models were considered significant with $p > 0.05$. Predictors within a model were considered significant with $p > 0.05$.

Statistical analyses for the repeated-measures ANOVAs that were performed on the behavioral data were performed using SPSS Statistics 27. All other statistical analyses were performed using customized code written in Matlab R2021b v.9.11.0 (see ***Data and Software Availability***).

Results

Modeling the relationship between Fractional Anisotropy and Age: Subfields undergo different developmental trajectories

To identify developmental trends in tissue microstructure within each hippocampal subfield, we segmented the hippocampus into the four hippocampal subfields (i.e., CA1, CA23, DG, SUB) and calculated the mean fractional anisotropy (FA) as a measure of tissue microstructure for each subfield. We then applied an explicit model testing approach to determine the best model for the relationship between FA, age, and sex. In each model, FA was entered as the response variable and age and sex were entered as predictors in one of seven possible configurations in order to test both linear and nonlinear models as well as main effect and interaction models. Pearson correlations among all variables are presented in **Supplementary Table 1**. Results of the model testing for each hippocampal subfield are displayed in **Table 2**.

FA in the CA1 and DG subfields were best modeled by a nonlinear model that included only a quadratic term for age, indicating a nonlinear main effect of age for both subfields and no effect of sex. For CA1, the best model was CA1 ~ age^2, *AICc* = 221.31, p = 0.006 (**Figure 2A**). For DG, the best model was DG ~ age^2, *AICc* = 210.00, p = 0.009 (**Figure 2C**).

FA in the CA23 and SUB subfield was best modeled by a non-linear interaction model that included sex and age. The best model was CA23 ~ sex*age^2, *AICc* = 203.16, p = 0.0001, indicating the presence of a nonlinear effect of age that differed between males and females (**Figure 2B**). For males, the nonlinear effect of age was parabolic increase; for females, the effect was an inverted U-shaped trajectory (**Figure 2B, inset**).

For SUB, the best model was SUB ~ age, *AICc* = 218.07, p = 0.025, indicating a linear main effect of age (**Figure 2D**). As age increased, FA decreased.

Of all the additional models tested in the supplementary analyses for the hippocampus and its subregions, we found no model that fit the data better than the models tested using the subfields. In fact, nearly all models were not significant. The only exception was a simple linear model with FA in the head subregion as the response variable and sex as the predictor variable, HEAD ~ sex, *AICc* = 255.03, p = 0.002. Results of these supplemental analyses can be found in **Supplementary Table 2**. Models tested with mean diffusivity (MD) as the dependent variable were similarly non-significant with the exception of TAIL ~ age, *AICc* = 105.41, p = 0.005, SUB ~ age, *AICc* = 121.91, p = 0.001. Results of the supplemental analyses for MD can be found in **Supplementary Table 3** for hippocampal subregions and in **Supplementary Table 4** for hippocampal subfields.

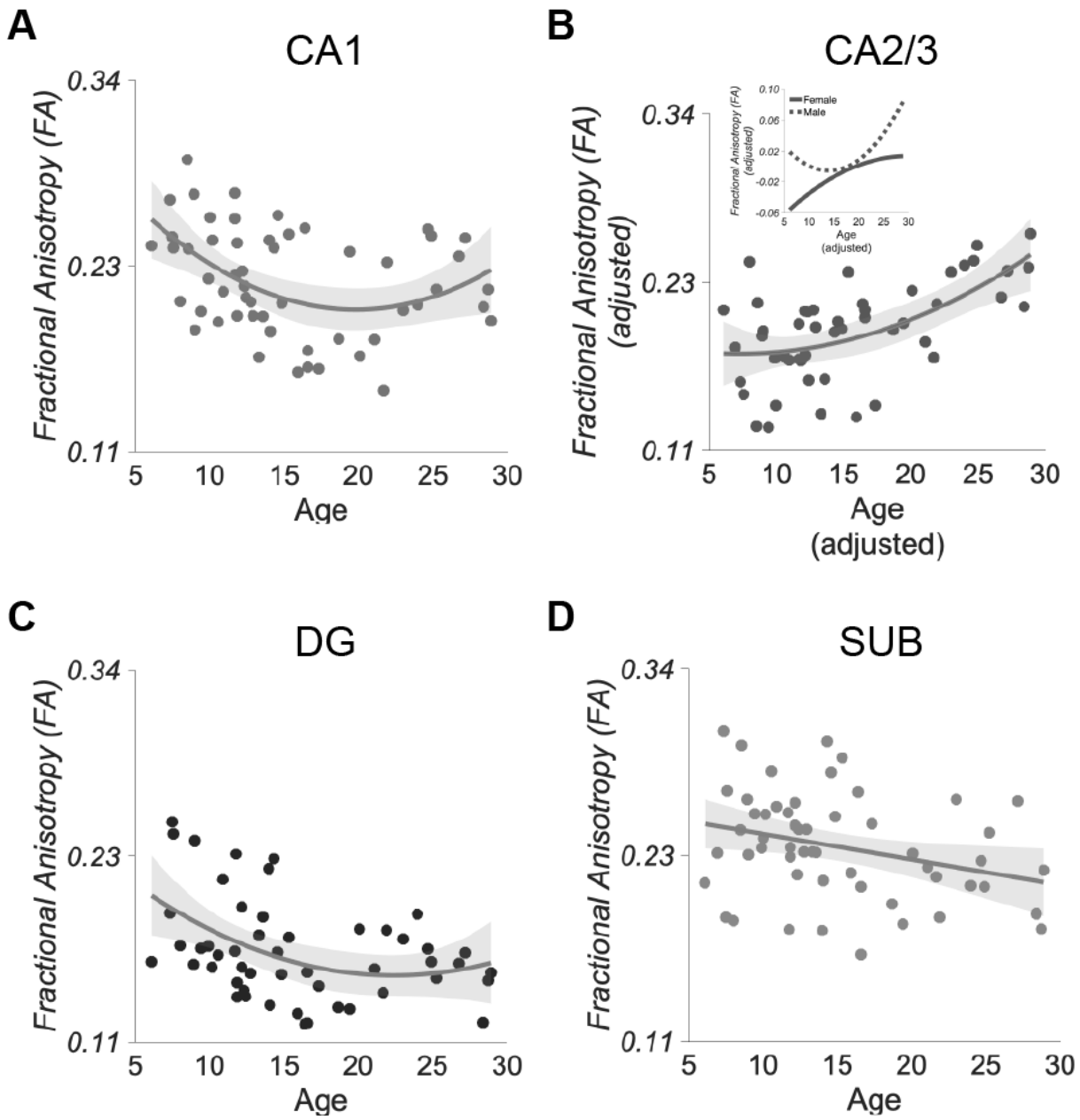

Figure 2. Different developmental trajectories for each subfield. Displayed are the best fitting models for the relationship between subfield fractional anisotropy (FA), age, and sex, selected based on *AICc*. **A.** *Cornu ammonis* field 1 (CA1). CA1 displayed a nonlinear, inverted-U shaped developmental trajectory. **B.** *Cornu ammonis* fields 2 and 3 (CA2/3). CA2/3 displayed a nonlinear interaction between sex and age, such that the nonlinear effect of age depended on sex. Males displayed a U-shaped trajectory while females displayed the opposite. **C.** Dentate Gyrus (DG). The DG displayed a nonlinear developmental trajectory, such that FA decreased with age. **D.** Subiculum (SUB). The SUB displayed a linear developmental trajectory, such that FA decreased with age.

Table 2. Models tested for Hippocampal Subfield and Age Relationship

Subfield	*Regression Model*	*N*	*AICc*	*F*	*p*
CA1	CA1 ~ age	54	219.35	2.00	0.163
	CA1 ~ age^2	**54**	**221.31**	**5.76**	**0.006**
	CA1 ~ sex	51	219.71	6.26	0.016
	CA1 ~ sex + age	52	218.16	2.57	0.087
	CA1 ~ sex + age^2	55	221.32	3.15	0.033
	CA1 ~ sex * age	54	217.25	1.92	0.139
	CA1 ~ sex * age^2	55	217.00	1.95	0.103
CA23	CA23 ~ age	50	197.30	7.49	0.117
	CA23 ~ age^2	51	196.62	4.20	0.021
	CA23 ~ sex	53	193.86	5.76	0.020
	CA23 ~ sex + age	49	201.37	10.20	0.0002
	CA23 ~ sex + age^2	49	200.83	7.46	0.0004
	CA23 ~ sex * age	49	198.99	6.64	0.0008
	CA23 ~ sex * age^2	**49**	**203.16**	**6.66**	**0.0001**
DG	DG ~ age	51	209.61	7.42	0.009
	DG ~ age^2	**50**	**210.00**	**5.27**	**0.009**
	DG ~ sex	51	205.87	0.012	0.913
	DG ~ sex + age	51	203.16	6.66	0.0001
	DG ~ sex + age^2	50	207.64	3.44	0.024
	DG ~ sex * age	51	205.05	2.39	0.081
	DG ~ sex * age^2	49	200.75	2.71	0.033
SUB	**SUB ~ age**	**54**	**218.07**	**5.30**	**0.025**
	SUB ~ age^2	54	215.87	2.62	0.083
	SUB ~ sex	55	216.92	1.59	0.213
	SUB ~ sex + age	54	217.64	3.56	0.036
	SUB ~ sex + age^2	54	215.47	2.39	0.080
	SUB ~ sex * age	53	216.31	2.66	0.059
	SUB ~ sex * age^2	52	213.41	2.51	0.043

Best fit models are bolded, were selected on AICc, and are displayed in Figure 2.

Memory-based inference improves over development

Memory-based inference: Matching and Inference

To investigate memory-based inference performance as a function of age, we performed a two-way repeated-measures ANOVA for Task (Matching, Inference) and Age Group (children, adolescents, adults). The dependent variable was either accuracy or reaction time. Additionally, we used a one-sample *t*-test to ensure that accuracy was above chance, i.e., greater than 33%. We note that a prior study reported results of this analysis (Schlichting et al. 2017), and are reproduced in the current study that uses an overlapping set of participants for the benefit of the reader.

Accuracy. Results revealed a greater difference between performance on the Matching task and the Inference task in children compared to adolescents (**Figure 3A**). All age groups performed above chance on both tasks, all *p*s < .001. A two-way repeated-measures ANOVA for Age Group (children, adolescents, adults) and Task (Matching, Inference) revealed a significant main effect of Age Group, $F(2, 655) = 8.900$, $p = 3.83\text{x}10^{-4}$, a significant main effect of Task, $F(1, 65) = 41.311$, $p = 1.79\text{x}10^{-8}$, and a significant interaction, $F(2, 65) = 4.388$, $p = 0.016$. There was a significant difference among the age groups on the Matching task, $F(2, 67) = 6.418$, $p = 0.003$,

and a significant difference among age groups on the Inference task, $F(2, 67) = 8.594$, $p = 4.88\text{x}10^{-4}$. For the Matching task, adult and adolescent performance were greater than child performance, $t(44) = 3.499$, $p = 0.001$ and $t(42) = 2.077$, $p = 0.044$, respectively, and there was no significant difference between adult and adolescent performance, $t(44) = 1.288$, $p = 0216$. For the Inference task, results were similar; adult and adolescent performance were greater than child performance, $t(44) = 3.999$, $p = 2.39\text{x}10^{-4}$ and $t(42) = 2.900$, $p = 0.006$, respectively, and there was no significant difference between adult and adolescent performance, $t(44) = 0.878$, $p = 0.385$. Performance on the Matching task was greater than performance the Inference task in children, $t(11) = 4.113$, $p = 0.002$, and in adolescents, $t(16) = 2.319$, $p = 0.034$, but not in adults, $t(21) = 1.962$, $p = 0.063$. The difference between Matching performance and Inference performance was greater in children than in adolescents, $t(42) = 2.320$, $p = 0.025$, and no different in adolescents than in adults, $t(44) = 0.365$, $p = 0.717$.

Reaction time. Results demonstrated that responses were slower on the Inference task than on the Matching task across age groups and that adolescents responded faster than adults and children across tasks (**Figure 3B**). A two-way repeated-measures ANOVA for Age Group (children, adolescents, adults) and Task (Matching, Inference) revealed a significant main effect of Age Group, $F(2, 65) = 4.236$, $p = 0.019$, and a main effect of Task, $F(1, 65) = 117.057$, $p = 3.54\text{x}10^{-16}$. The interaction did not reach significance, $F(2, 65) = 1.1778$, $p = 0.177$. Adolescents responded faster than children, $t(42) = 2.861$, $p = 0.007$, and faster than adults, $t(44) = 2.003$, $p = 0.051$. The difference between children and adults was not significant, $t(44) = 1.155$, $p = 0.254$. Responses were faster on the Matching task compared to the Inference task, $t(67) = 10.779$, $p = 2.81\text{x}10^{-16}$.

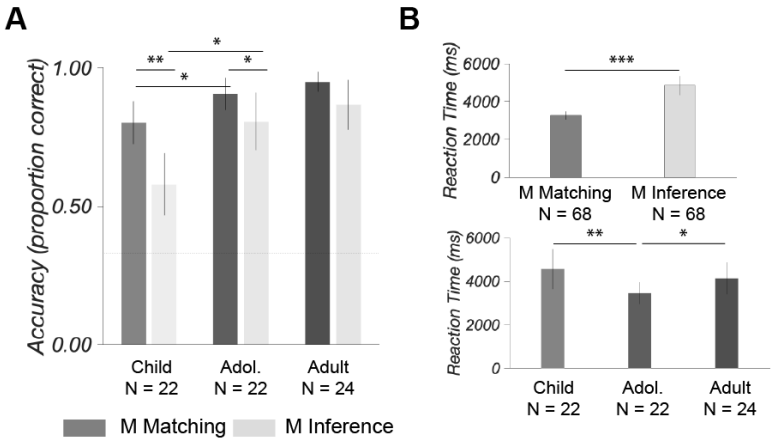

Figure 3. Memory-based inference: Matching and Inference. A. Accuracy. Performance was above chance on both tasks; chance performance was 33% (dashed line). Performance on the Matching task was greater than performance on the Inference task within the child and adolescent age groups. The difference between Matching and Inference performance was greater in children compared to adolescents. Note: This is a replication of Schlichting et al. (2017), Figure 4A. **B.** Reaction Time. Reaction time was slower on the Inference task than on the Matching task across age groups (top) and slower in adolescents than in children and adults across tasks (bottom). Error bars are 95% confidence intervals. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Perception-based inference: Matching and Inference

To investigate perception-based inference performance as a function of age, we performed a two-way repeated-measures ANOVA for Task (Matching, Inference) and Age Group (children, adolescents, adults). The dependent variable was either accuracy or reaction time. Additionally, we used a one-sample *t*-test to ensure that accuracy was above chance, i.e., greater than 50%.

Accuracy. Results demonstrated that performance was greater on the P Matching task than on the Inference task across age groups (**Figure 4A**). All age groups performed above chance on the Matching and inference tasks, all *p*s < .001. A two-way repeated-measures ANOVA for Age Group (children, adolescents, adults) and Task (Matching, Inference) revealed a significant main effect of Task, $F(1, 63) = 73.497$, $p = 3.57\text{x}10^{-12}$. The main effect of Age Group did not reach significance, $F(2, 63) = 2.791$, $p = 0.069$. The interaction was not significant, $F(2, 63) = 0.918$, $p = 0.405$. Performance was greater on the Matching task than on the Inference task, $t(65) = 8.151$, $p = p = 3.57\text{x}10^{-12}$.

Reaction time. Results demonstrated that children responded slower than adolescents and adults on the Matching task but not on the Inference task (**Figure 4B**). A two-way repeated-measures ANOVA for Age Group (children, adolescents, adults) and Task (Matching, Inference) revealed a significant main effect of Task, $F(1, 63) = 316.511$, $p = 2.97\text{x}10^{-26}$, and a significant interaction, $F(2, 63) = 4.746$, $p = 0.012$. The main effect of Age Group was not significant, $F(2, 63) =1.091$, $p = 0.342$. Responses were slower in the Inference task than in the Matching task, $t(65) = 16.946$, $p = 1.54\text{x}10^{-25}$. There was a significant difference among age groups for the Matching task, $F(2, 65) = 5.615$, $p = 0.006$, that was not present for the Inference task, $F(2, 65) = 1.249$, $p = 0.294$. On the Matching task, children responded slower than adolescents, $t(42) = 2.370$, $p = 0.022$, and than adults, $t(40) = 2.739$, $p = 0.009$, and there was no difference between adolescents and adults, $t(44) = 0.734$, $p = 0.467$.

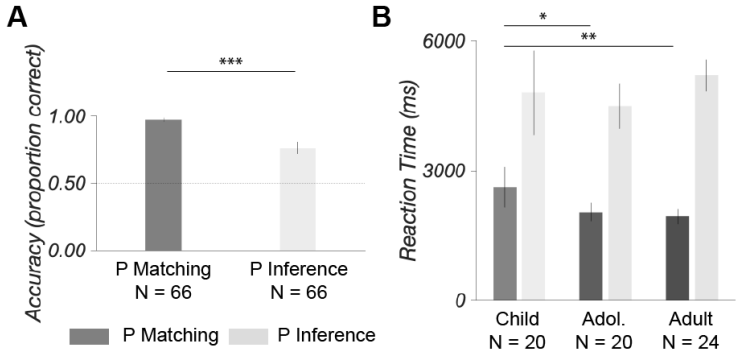

Figure 4. Perception-based inference: Matching and Inference. **A.** Accuracy. Performance was above chance on both tasks; chance performance was 50% (dashed line). Performance on the Inference task was less than performance on the Matching task. **B.** Reaction Time. On the Matching task, children responded slower than adolescents and adults. Error bars are 95% confidence intervals. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Microstructure of hippocampal subfields related to associative inference

We assessed the relationship between the microstructure of hippocampal subfields and inference behavior when memory was required and when memory was not required. The goal was to explore the extent to which the cellular organization within hippocampal subfields might be related to inference behaviors by manipulating the memory requirement during inference. Associative inference requires memory and develops gradually, suggesting that the cellular computations performed within hippocampal subfields likely track associative inference ability and not perception-based inference ability. We constructed a separate regression model for each subfield that contained a predictor for each task of interest. One model included predictors for the inference tasks: M Inference and P Inference. The second model included predictors for the matching tasks: M Matching and P Matching. Both models included covariates of no interest, including IQ and the specific sex and age predictors that explained the most variance in subfield microstructure in the earlier analyses (see **Modelling to determine developmental trajectory: Subfields undergo different developmental trajectories**). The response variable was always the mean microstructural measurement (i.e., FA) for that subfield, averaged over all voxels included in the ROI for that subfield.

Cornu ammonis field 1, CA1

Inference tasks

There was no statistically significant relationship between either of the inference tasks and CA1 microstructure. Multiple linear regression was used to test if performance on either of the inference tasks significantly predicted CA1 microstructure. The fitted regression model was CA1 ~ M Inference + P Inference + iq + age^2. The overall regression was statistically significant, R^2 = 0.214, *Adj.* R^2 = 0.132, *AICc* = 215.98, $F(5, 48) = 2.61$, $p = 0.036$. P Matching significantly predicted CA1 microstructure, $\beta = 0.222$, $p = 0.044$, as did age, $\beta = 0.0004$, $p = 0.008$, and age^2, $\beta = -0.013$, $p = 0.004$. There were no other significant predictors, all *p*s > 0.05.

Additionally, we fitted a regression model that tested for age-related changes in the relationship between CA1 microstructure and task performance, CA1 ~ M Inference*age + P Inference*age + iq, but the overall regression was not statistically significant, $R^2 = 0.077$, *Adj.* $R^2 = 0.041$, *AICc* = 209.44, $F(6, 47) = 0.652$, $p = 0.688$, and there were no significant predictors, all *p*s > 0.05.

Matching control tasks

Performance on the perception-based Matching task significantly predicted CA1 microstructure (**Figure 5A**), suggesting that more accurate performance on perception-based pair matching was related to greater directional coherence of cellular tissue within the CA1 subfield. Multiple linear regression was used to test if performance on any of the matching tasks significantly predicted CA1 microstructure. The fitted regression model was CA1 ~ M Matching + P Matching + iq + age^2. The overall regression was statistically significant, $R^2 = 0.223$, *Adj.* $R^2 = 0.141$, *AICc* = 217.05, $F(5, 47) = 2.70$, $p = 0.032$. Age significantly predicted CA1 microstructure, $\beta = -0.013$,

p = 0.007, as did age^2, β = 0.0003, p = 0.012. There were no other significant predictors, all ps > 0.05.

Additionally, we fitted a regression model that tested for age-related changes in the relationship between CA1 microstructure and task performance, CA1 ~ M Matching*age + P Matching*age + iq, but the overall regression was not statistically significant, R^2 = 0.083, *Adj.* R^2 = 0.034, *AICc* = 209.78, F(6, 47) = 0.709, p = 0.644, and there were no significant predictors, all ps > 0.05.

Cornu ammonis fields 2 and 3, CA2/3

Inference tasks

There was no statistically significant relationship between either of the inference tasks and CA2/3 microstructure. Multiple linear regression was used to test if performance on any of the inference tasks significantly predicted CA2/3 microstructure. The fitted regression model was CA2/3 ~ M Inference + P Inference + iq + sex*age^2. The overall regression was statistically significant, R^2 = 0.527, *Adj.* R^2 = 0.430, *AICc* = 199.35, F(8, 39) = 5.44, p = 0.0001. IQ significantly predicted CA2/3 microstructure, β = 0.0008, p = 0.035, as did sex, β = 0.177, p = 0.004, sex*age, β = -0.019, p = 0.016, and sex*age^2, β = 0.0005, p = 0.019. There were no other significant predictors, all ps > 0.05.

Additionally, we fitted a regression model that tested for age-related changes in the relationship between CA2/3 microstructure and task performance, CA23 ~ M Inference*age + P Inference*age + iq. The overall regression was statistically significant, R^2 = 0.245, *Adj.* R^2 = 0.140, *AICc* = 101.71, F(6, 43) = 2.33, p = 0.049. P Inference performance significantly predicted CA2/3 microstructure, β = -0.112, p = 0.041. There were no other significant predictors, all ps > 0.05.

Matching control tasks

There was no statistically significant relationship between any of the behavioral tasks and CA2/3 microstructure. Multiple linear regression was used to test if performance on any of the learning tasks significantly predicted CA2/3 microstructure. The fitted regression model was CA2/3 ~ M Matching + P Matching + iq + sex*age^2. The overall regression was statistically significant, R^2 = 0.437, *Adj.* R^2 = 0.329, *AICc* = 193.17, F(8, 42) = 4.07, p = 0.001. IQ significantly predicted CA2/3 microstructure, β = 0.001, p = 0.021, as did sex, β = 0.183, p = 0.012, and sex*age, β = -0.018, p = 0.051. There were no other significant predictors, all ps > 0.05.

Additionally, we fitted a regression model that tested for age-related changes in the relationship between CA2/3 microstructure and task performance, CA23 ~ M Matching*age + P Matching*age + iq. The overall regression did not reach statistical significance, R^2 = 0.146, *Adj.* R^2 = 0.032, *AICc* = 184.97, F(6, 45) = 1.28, p = 0.284, and there were no significant predictors, all ps > 0.05.

Dentate gyrus, DG

Inference tasks

Performance on the M Inference task significantly predicted DG microstructure, suggesting that the DG is related to inference that requires memory (**Figure 5C**). However, we again found no interaction with age, suggesting that the association between dentate gyrus microstructure and inference tasks may not vary much with age.

Multiple linear regression was used to test if performance on any of the inference tasks significantly predicted DG microstructure. The fitted regression model was DG ~ M Inference + P Inference + iq + age^2. The overall regression was statistically significant, R^2 = 0.420, *Adj.* R^2 = 0.353, *AICc* = 213.64, $F(5, 43) = 6.24$, $p = 0.0002$. M Inference performance significantly predicted DG microstructure, $\beta = -0.036$, $p = 0.030$, as did IQ, $\beta = 0.0011$, $p = 0.0005$. There were no other significant predictors, all *p*s > 0.05.

Additionally, we fitted a regression model that tested for age-related changes in the relationship between DG microstructure and task performance, DG ~ M Inference*age + P Inference*age + iq. The overall regression was statistically significant, R^2 = 0.406, *Adj.* R^2 = 0.323, *AICc* = 210.53, $F(6, 43) = 4.89$, $p = 0.0007$. M Inference performance significantly predicted DG microstructure, $\beta = -0.082$, $p = 0.040$, as did IQ, $\beta = 0.001$, $p = 0.002$. There were no other significant predictors, all *p*s > 0.05.

Matching control tasks

There was no statistically significant relationship between any of the behavioral tasks and DG microstructure, suggesting that the association between dentate microstructure and matching may not vary much with age.

Multiple linear regression was used to test if performance on any of the matching tasks significantly predicted DG microstructure. The fitted regression model was DG ~ M Matching + P Matching + iq + age^2. The overall regression was statistically significant, R^2 = 0.251, *Adj.* R^2 = 0.166, *AICc* = 206.89, $F(5, 44) = 2.95$, $p = 0.022$; however, there were no significant predictors, all *p*s > 0.05.

Additionally, we fitted a regression model that tested for age-related changes in the relationship between DG microstructure and task performance, DG ~ M Matching*age + P Matching*age + iq. The overall regression did not reach statistical significance, R^2 = 0.217, *Adj.* R^2 = 0.105, *AICc* = 203.78, $F(6, 42) = 1.94$, $p = 0.096$, and there were no significant predictors, all *p*s > 0.05.

Subiculum, SUB

Inference tasks

Performance on the P Inference task significantly predicted SUB microstructure, suggesting that the SUB is related to inference that does not require memory (**Figure 5B**). Multiple linear regression was used to test if performance on any of the inference tasks significantly predicted SUB microstructure. The fitted regression model was SUB ~ M Inference + P Inference + iq + age. The overall regression was statistically significant, R^2 = 0.322, *Adj.* R^2 = 0.260, *AICc* = 216.70, $F(4, 44) = 5.21$, $p = 0.002$. P Inference performance significantly predicted SUB microstructure, $\beta = 0.050$, $p = 0.003$, as did age, $\beta = -0.002$, $p = 0.014$. There were no other significant predictors, all *p*s > 0.05.

Additionally, we fitted a regression model that tested for age-related changes in the relationship between SUB microstructure and task performance, SUB ~ M Inference*age + P Inference*age + iq. The overall regression did not reach statistical significance, R^2 = 0.228, *Adj.* R^2 = 0.120, *AICc* = 212.50, $F(6, 43) = 2.12$, $p = 0.071$, and there were no significant predictors, all *p*s > 0.05.

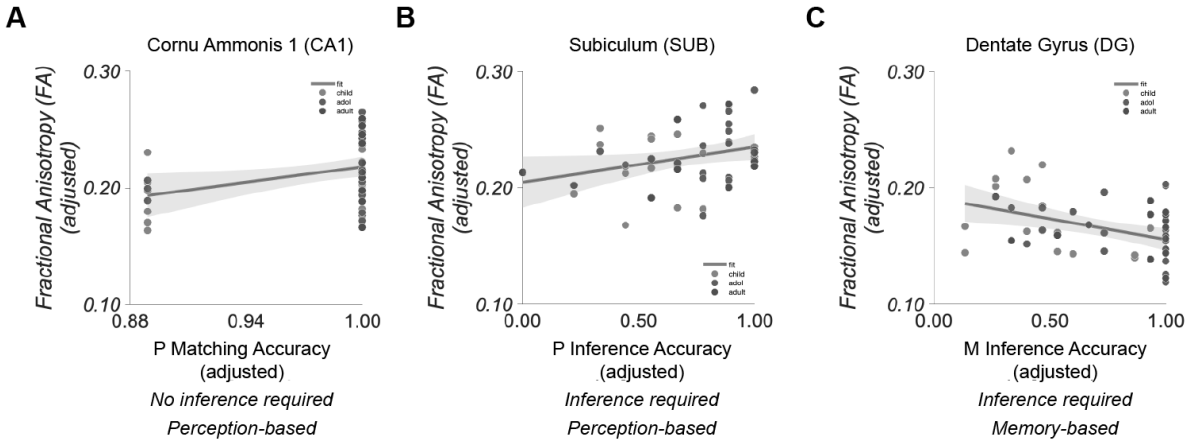

Figure 5. Relationship Between Subfield Microstructure and Associative Inference Performance. A. Perception-based matching: P Matching. P Matching accuracy significantly predicted CA1 microstructure. **B.** Perception-based inference: P Inference. P Inference accuracy significantly predicted SUB microstructure. **C.** Memory-based inference: M Inference. M Inference accuracy significantly predicted DG microstructure. Each plot describes the relationship between the fitted response as a function of accuracy with the other predictor in the model averaged out.

Matching control tasks

There was no statistically significant relationship between any of the behavioral tasks and SUB microstructure. Multiple linear regression was used to test if performance on any of the matching tasks significantly predicted SUB microstructure. The fitted regression model was SUB ~ M Matching + P Matching + iq + age. The overall regression did not reach statistical significance, R^2 = 0.143, Adj. R^2 = 0.069, AICc = 209.14, $F(4, 46) = 1.93$, $p = 0.122$, yet one predictor reached statistical significance, age, $\beta = -0.002$, $p = 0.030$. There were no other significant predictors, all ps > 0.05.

Additionally, we fitted a regression model that tested for age-related changes in the relationship between SUB microstructure and task performance, SUB ~ M Matching*age + P Matching*age + iq. The overall regression was statistically significant, $R^2 = 0.350$, Adj. $R^2 = 0.254$, AICc = 213.78, $F(6, 41) = 3.67$, $p = 0.005$. M Matching performance significantly predicted SUB microstructure, $\beta = 1.334$, $p = 0.010$, as did age*ABMatching, $\beta = -0.012$, $p = 0.002$, and IQ, $\beta = 0.001$, $p = 0.032$. There were no other significant predictors, all ps > 0.05.

Discussion

Our goal was to determine the developmental trajectories of the tissue properties of hippocampal subfields and their relationships to the development of associative inference and memory. We characterized the relationship between age and the cellular organization within hippocampal subfields and found that subfields have distinct developmental trajectories, with a linear trajectory for the SUB and non-linear trajectories for the CA1, CA2/3, and DG that depended on sex in the CA2/3 subfield. We found that the tissue properties of the DG and SUB were both related to associative inference performance, independent of age, with the DG being related to associative inference that required memory and the SUB being related to associative inference that did not require memory. Overall, our results suggest that hippocampal subfields undergo distinct developmental trajectories in terms of cellular organization and that the cellular organization of hippocampal subfields have distinct relationships with associative inference and memory.

Microstructures of hippocampal subfields undergo different developmental trajectories

We observed age-related changes in hippocampal subfields, finding that different subfields have unique developmental trajectories from childhood to adulthood. While CA1 and DG subfields displayed a non-linear U-shaped trajectory, with the lowest FA in adolescents for CA1 and in adulthood for DG, the SUB displayed a linear relationship with age, with FA decreasing from childhood to adulthood. The CA2/3 subfield displayed a more complicated trajectory, with males undergoing a U-shaped trajectory with the lowest FA in adolescents, and females undergoing an inverted U-shaped trajectory with the lowest FA in childhood.

The development of cellular organization within the hippocampus has been investigated using measures of volume and mean diffusivity across the entire hippocampus, hippocampal subregions, and hippocampal subfields (Schlichting et al. 2017; Riggins et al. 2015; Daugherty, Flinn, and Ofen 2017; Lee, Ekstrom, and Ghetti 2014; Pereira et al. 2014; Hamer et al. 2008; Giedd et al. 1996; Ostby et al. 2009; Lee, Johnson, and Ghetti 2017; Tamnes et al. 2018; Callow, Canada, and Riggins 2020; Wolf et al. 2015; Mah, Geeraert, and Lebel 2017). In the present work we employed an additional metric, fractional anisotropy (FA), to assess cellular organization within the hippocampus. We selected this metric based on results from animal models that suggested that subfields have distinct cellular organizations that could be captured with FA (Sierra et al. 2015; Salo et al. 2017) and that FA and MD provide independent information in human hippocampus (Treit et al. 2018; Shereen et al. 2011). Our results suggest that distinct developmental trajectories for hippocampal subfields can be measured using FA. Notably, we did not find age-related changes in FA within hippocampal subregions (see **Supplementary materials**), suggesting that the developmental unit of hippocampal FA may be the subfield as opposed to other common segmentations.

At least one prior work has reported age-related changes in microstructure of hippocampal subregions (Langnes et al. 2020). Using a cross-sectional design across the ages of 4 to 93

years, this study reported different developmental trajectories for anterior and posterior hippocampus: MD in anterior hippocampus decreases from childhood to approximately 35 years of age while MD in posterior hippocampus remains relatively stable. After the age of 35 years, MD in both anterior and posterior hippocampus increases dramaticaly. In the current study, we segmented the hippocampus into the head, body, and tail but found no evidence of age-related changes in these subregions (see **Supplemental Materials**). However, results from the FA analysis within each subfield are consistent with different developmental trends for anterior and posterior hippocampus, but suggest that age-related changes detected in the larger anterior posterior subregions may be driven by subfield-selective changes in cellular organization. More research will be necessary to better understand the relationship between developmental trends in cellular organization detected in larger subregion segmentations and trends detected in smaller subfield segmentations.

Associative inference performance that requires memory improves over development

Associative inference improves gradually throughout development, similar to memory processing more broadly (Andrews & Halford, 1998; Bauer, Varga, King, Nolen, & White, 2015; Halford, 1984; Townsend, Richmond, Vogel-Farley, Thomas, 2010; Ghetti & Bunge, 2012). Our results suggest that, although associative inference and memory both develop gradually, they do not share similar developmental trajectories. We found that performance on an associative inference task that did not require memory was no different among children, adolescents, and adults, suggesting that associative inference in the absence of a memory requirement may reach adult-like performance earlier than 6 years of age. In the same participant sample, performance on an associative inference task that did require memory was greater in adults compared to adolescent children and greater in adolescent children compared to younger children (originally reported in Schlichting et al., 2017), similar to the prolonged trajectory expected for associative inference and memory. Different trajectories for associative inference with and without memory would suggest that associative inference and memory are separate constructs, at least during development.

Microstructures of hippocampal subfields related to associative inference and memory

Understanding the relationship between subfield development, associative inference, and memory in terms of subfield tissue microstructure can provide insight into the nature of memory and the neural mechanisms supporting memory. We found that DG microstructure was related to associative inference performance when memory was required, but that SUB microstructure was related to associative inference performance when memory was not required. Furthermore, CA1 microstructure was related to pair matching that required memory. Notably, the relationships among subfield microstructure and behavior were independent of age, suggesting that age may not be a driving factor in the relationship between subfield cellular organization and associative inference and memory.

One theory of memory proposes that the hippocampus proactively integrates new memories with existing memories through a process of overlapping memory traces (Schlichting and

Preston 2015). In adults, the CA1 subfield is thought to use novelty as a cue to trigger memory integration. If input from the environment is inconsistent with predictions based on prior memory traces, CA1 influences the plasticity of the CA3 subfield so as to assist in the binding of the novel input with associated prior memories (McKenzie et al. 2014). The dentate gyrus (DG) is thought to assist CA3 in memory integration by performing pattern separation, a crucial process that prevents interference among inputs through sparse connections and high inhibitory processes (Marr 1971; Schapiro et al. 2017). Thus, while CA1/CA3 may be focused on binding shared features across episodes, DG/CA3 may be more focused on separating features that are not shared. The function of the subiculum (SUB) is less clear, however it is more densely connected to regions within and outside of the hippocampus than subfields in the hippocampus proper (i.e., CA1, CA2/3, DG), suggesting that it has functions that are potentially independent of core hippocampal functions (Aggleton and Christiansen 2015).

Our results suggest that hippocampal subfields may play important, yet distinct, roles in integrating new memories with existing memories. We observed different behavioral correlates of subfield cellular organization, namely that CA1 was related to pair matching while DG was related to associative inference that required memory. Both CA1 and DG assist in memory integration by altering the cellular organization and plasticity of the CA3 subfield. Although more work is needed to clarify the implications of our results, we suggest that our results support the proposal that CA1 and DG have unique functional properties that interface with neural mechanisms present in the CA3 subfield to support memory encoding. While CA1 may assist CA3 in detecting novelty, it may not be sufficient for memory encoding without concurrent assistance from DG to perform pattern separation, leaving CA1 less related to memory outcomes than DG.

Our results are consistent with the notion that the SUB may have functions that are independent of core hippocampal functions (Aggleton and Christiansen 2015). Results suggest that the cellular organization with the SUB was related to associative inference and not necessarily memory. It is possible that the SUB interfaces with other cortical mechanisms to support reasoning in the absence of relevant memories, given that it is more densely connected to regions outside of the hippocampus than the other subfields. More research will be necessary to better understand the role of the SUB in associative inference and memory.

Conclusion

The cellular organization of hippocampal subfields undergo different developmental trajectories and exhibit distinct relations with associative inference and memory.

Acknowledgements

Sophia Vinci-Booher was supported by an NSF SBE Postdoctoral Fellowship (SMA-2004877). Franco Pestilli was supported with grants from the National Science Foundation (OAC-1916518, IIS-1912270, IIS-1636893, BCS-1734853), National Institutes of Health (NIMH R01MH126699, NIBIB R01EB030896, NIBIB R01EB029272), and Wellcome Trust (226486/Z/22/Z), as well as a Microsoft Investigator Fellowship, and a gift from the Kavli Foundation.

References

Achenbach, and Edelbrock. 1991. “Child Behavior Checklist.” *Burlington (Vt)*. https://www.academia.edu/download/47232246/Assessing__Childrens__Well-Being.pdf#page=86.

Ades-Aron, Benjamin, Jelle Veraart, Peter Kochunov, Stephen McGuire, Paul Sherman, Elias Kellner, Dmitry S. Novikov, and Els Fieremans. 2018. “Evaluation of the Accuracy and Precision of the Diffusion Parameter EStImation with Gibbs and NoisE Removal Pipeline.” *NeuroImage* 183 (December): 532–43.

Aggleton, John P., and Kat Christiansen. 2015. “Chapter 4 - The Subiculum: The Heart of the Extended Hippocampal System.” In *Progress in Brain Research*, edited by Shane O’Mara and Marian Tsanov, 219:65–82. Elsevier.

Assaf, Yaniv, and Ofer Pasternak. 2008. “Diffusion Tensor Imaging (DTI)-Based White Matter Mapping in Brain Research: A Review.” *Journal of Molecular Neuroscience: MN* 34 (1): 51–61.

Avants, Brian B., Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein, and James C. Gee. 2011. “A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration.” *NeuroImage* 54 (3): 2033–44.

Barron, Helen C., Raymond J. Dolan, and Timothy E. J. Behrens. 2013. “Online Evaluation of Novel Choices by Simultaneous Representation of Multiple Memories.” *Nature Neuroscience* 16 (10): 1492–98.

Basser, Peter J., and Derek K. Jones. 2002. “Diffusion-Tensor MRI: Theory, Experimental Design and Data Analysis - a Technical Review.” *NMR in Biomedicine* 15 (7-8): 456–67.

Basser, P. J., J. Mattiello, and D. LeBihan. 1994. “MR Diffusion Tensor Spectroscopy and Imaging.” *Biophysical Journal* 66 (1): 259–67.

Bechara, A., A. R. Damasio, H. Damasio, and S. W. Anderson. 1994. “Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex.” *Cognition* 50 (1-3): 7–15.

Brainerd, C. J., and V. F. Reyna. 1992. “Explaining ‘Memory Free’ Reasoning.” *Psychological Science* 3 (6): 332–39.

Callow, Daniel D., Kelsey L. Canada, and Tracy Riggins. 2020. “Microstructural Integrity of the Hippocampus During Childhood: Relations With Age and Source Memory.” *Frontiers in Psychology* 11 (September): 568953.

Carlesimo, Giovanni A., Andrea Cherubini, Carlo Caltagirone, and Gianfranco Spalletta. 2010. “Hippocampal Mean Diffusivity and Memory in Healthy Elderly Individuals: A Cross-Sectional Study.” *Neurology* 74 (3): 194–200.

Cattell, R. B. 1987. *Intelligence: Its Structure, Growth and Action*. Elsevier.

Crone, Eveline A., Carter Wendelken, Linda van Leijenhorst, Ryan D. Honomichl, Kalina Christoff, and Silvia A. Bunge. 2009. “Neurocognitive Development of Relational Reasoning.” *Developmental Science* 12 (1): 55–66.

Datson, N. A., L. Meijer, P. J. Steenbergen, M. C. Morsink, S. van der Laan, O. C. Meijer, and E. R. de Kloet. 2004. “Expression Profiling in Laser-Microdissected Hippocampal Subregions in Rat Brain Reveals Large Subregion-Specific Differences in Expression.” *The European Journal of Neuroscience* 20 (10): 2541–54.

Daugherty, Ana M., Robert Flinn, and Noa Ofen. 2017. “Hippocampal CA3-Dentate Gyrus Volume Uniquely Linked to Improvement in Associative Memory from Childhood to Adulthood.” *NeuroImage* 153 (June): 75–85.

Derogatis, L. R. 1977. *SCL-90-R: Administration, Scoring, and Procedures Manual*. Baltimore: Clinical Psychometric Research.

Destrieux, Christophe, Bruce Fischl, Anders Dale, and Eric Halgren. 2010. “Automatic Parcellation of Human Cortical Gyri and Sulci Using Standard Anatomical Nomenclature.” *NeuroImage* 53 (1): 1–15.

Fischl, Bruce. 2012. “FreeSurfer.” *NeuroImage* 62 (2): 774–81.

Fiser, J., and R. N. Aslin. 2001. “Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes.” *Psychological Science* 12 (6): 499–504.

Fonov, Vladimir, Alan C. Evans, Kelly Botteron, C. Robert Almli, Robert C. McKinstry, D. Louis Collins, and Brain Development Cooperative Group. 2011. “Unbiased Average Age-Appropriate Atlases for Pediatric Studies.” *NeuroImage* 54 (1): 313–27.

Frigge, Michael, David C. Hoaglin, and Boris Iglewicz. 1989. “Some Implementations of the Boxplot.” *The American Statistician* 43 (1): 50–54.

Giedd, J. N., A. C. Vaituzis, S. D. Hamburger, N. Lange, J. C. Rajapakse, D. Kaysen, Y. C. Vauss, and J. L. Rapoport. 1996. “Quantitative MRI of the Temporal Lobe, Amygdala, and Hippocampus in Normal Human Development: Ages 4-18 Years.” *The Journal of Comparative Neurology* 366 (2): 223–30.

Glasser, Matthew F., Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, et al. 2013. “The Minimal Preprocessing Pipelines for the Human Connectome Project.” *NeuroImage* 80 (October): 105–24.

Greve, Douglas N., and Bruce Fischl. 2009. “Accurate and Robust Brain Image Alignment Using Boundary-Based Registration.” *NeuroImage* 48 (1): 63–72.

Halford, Joan Montgomery. 1998. “Easing the Way for Teachers.” *Educational Leadership: Journal of the Department of Supervision and Curriculum Development, N.E.A* 55 (5): 33–36.

Hamer, Lin, Gerig, and Gilmore. 2008. “A Structural MRI Study of Human Brain Development from Birth to 2 Years.” *Journal of Atomic and Molecular Physics*. https://www.jneurosci.org/content/28/47/12176.short.

Hayashi, Soichi, Lindsey Kitchell, and Franco Pestilli. 2017. *Freesurfer*. brainlife.io. https://doi.org/10.25663/BL.APP.0.

Hayashi, Soichi, Brent McPherson, and Brad Caron. 2018. *HCP ACPC Alignment (T1)*. brainlife.io. https://doi.org/10.25663/BL.APP.99.

Hirshhorn, Marnie, Cheryl Grady, R. Shayna Rosenbaum, Gordon Winocur, and Morris Moscovitch. 2012. “The Hippocampus Is Involved in Mental Navigation for a Recently Learned, but Not a Highly Familiar Environment: A Longitudinal fMRI Study.” *Hippocampus* 22 (4): 842–52.

Hsu, Nina S., Margaret L. Schlichting, and Sharon L. Thompson-Schill. 2014. “Feature Diagnosticity Affects Representations of Novel and Familiar Objects.” *Journal of Cognitive Neuroscience* 26 (12): 2735–49.

Hurvich, Clifford M., and Chih-Ling Tsai. 1991. “Bias of the Corrected AIC Criterion for Underfitted Regression and Time Series Models.” *Biometrika* 78 (3): 499–509.

Iaria, Giuseppe, Linda J. Lanyon, Christopher J. Fox, Deborah Giaschi, and Jason J. S. Barton. 2008. “Navigational Skills Correlate with Hippocampal Fractional Anisotropy in Humans.” *Hippocampus* 18 (4): 335–39.

Langnes, Espen, Markus H. Sneve, Donatas Sederevicius, Inge K. Amlien, Kristine B. Walhovd, and Anders M. Fjell. 2020. “Anterior and Posterior Hippocampus Macro- and Microstructure across the Lifespan in Relation to Memory-A Longitudinal Study.” *Hippocampus* 30 (7): 678–92.

Lavenex, Pierre, Pamela Banta Lavenex, and David G. Amaral. 2007. “Postnatal Development of the Primate Hippocampal Formation.” *Developmental Neuroscience* 29 (1-2): 179–92.

Le Bihan, Denis. 2014. “Diffusion MRI: What Water Tells Us about the Brain.” *EMBO Molecular Medicine* 6 (5): 569–73.

Lee, Joshua K., Arne D. Ekstrom, and Simona Ghetti. 2014. “Volume of Hippocampal Subfields and Episodic Memory in Childhood and Adolescence.” *NeuroImage* 94 (July): 162–71.

Lee, Joshua K., Elliott G. Johnson, and Simona Ghetti. 2017. “Hippocampal Development: Structure, Function and Implications.” In *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition*, edited by Deborah E. Hannula and Melissa C. Duff, 141–66. Cham: Springer International Publishing.

Lister, James P., and Carol A. Barnes. 2009. “Neurobiological Changes in the Hippocampus during Normative Aging.” *Archives of Neurology* 66 (7): 829–33.

Mah, Alyssa, Bryce Geeraert, and Catherine Lebel. 2017. “Detailing Neuroanatomical Development in Late Childhood and Early Adolescence Using NODDI.” *PloS One* 12 (8): e0182340.

Manns, Joseph R., and Howard Eichenbaum. 2006. “Evolution of Declarative Memory.” *Hippocampus* 16 (9): 795–808.

Marr, D. 1971. “Simple Memory: A Theory for Archicortex.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 262 (841): 23–81.

McKenzie, Sam, Andrea J. Frank, Nathaniel R. Kinsky, Blake Porter, Pamela D. Rivière, and Howard Eichenbaum. 2014. “Hippocampal Representation of Related and Opposing Memories Develop within Distinct, Hierarchically Organized Neural Schemas.” *Neuron* 83 (1): 202–15.

McPherson, Brent. 2018. *mrtrix3 Preprocess*. brainlife.io. https://doi.org/10.25663/BL.APP.68.

Müller, M. J., M. Mazanek, C. Weibrich, P. R. Dellani, P. Stoeter, and A. Fellgiebel. 2006. “Distribution Characteristics, Reproducibility, and Precision of Region of Interest-Based Hippocampal Diffusion Tensor Imaging Measures.” *AJNR. American Journal of Neuroradiology* 27 (2): 440–46.

Munnelly, Anita, and Simon Dymond. 2014. “Relational Memory Generalization and Integration in a Transitive Inference Task with and without Instructed Awareness.” *Neurobiology of Learning and Memory* 109 (March): 169–77.

Ostby, Ylva, Christian K. Tamnes, Anders M. Fjell, Lars T. Westlye, Paulina Due-Tønnessen, and Kristine B. Walhovd. 2009. “Heterogeneity in Subcortical Brain Development: A Structural Magnetic Resonance Imaging Study of Brain Maturation from 8 to 30 Years.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 29 (38): 11772–82.

Pereira, Joana B., Cinta Valls-Pedret, Emilio Ros, Eva Palacios, Carles Falcón, Nuria Bargalló, David Bartrés-Faz, Lars-Olof Wahlund, Eric Westman, and Carme Junque. 2014. “Regional Vulnerability of Hippocampal Subfields to Aging Measured by Structural and Diffusion MRI.” *Hippocampus* 24 (4): 403–14.

Pierpaoli, C., and P. J. Basser. 1996. “Toward a Quantitative Assessment of Diffusion Anisotropy.” *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 36 (6): 893–906.

Preston, Alison R., Yael Shrager, Nicole M. Dudukovic, and John D. E. Gabrieli. 2004. “Hippocampal Contribution to the Novel Use of Relational Information in Declarative Memory.” *Hippocampus* 14 (2): 148–52.

Richland, Lindsey E., Robert G. Morrison, and Keith J. Holyoak. 2006. “Children’s Development of Analogical Reasoning: Insights from Scene Analogy Problems.” *Journal of Experimental Child Psychology* 94 (3): 249–73.

Riggins, Tracy, Sarah L. Blankenship, Elizabeth Mulligan, Katherine Rice, and Elizabeth

Redcay. 2015. “Developmental Differences in Relations Between Episodic Memory and Hippocampal Subregion Volume During Early Childhood.” *Child Development* 86 (6): 1710–18.

Salo, Raimo A., Tuukka Miettinen, Teemu Laitinen, Olli Gröhn, and Alejandra Sierra. 2017. “Diffusion Tensor MRI Shows Progressive Changes in the Hippocampus and Dentate Gyrus after Status Epilepticus in Rat - Histological Validation with Fourier-Based Analysis.” *NeuroImage* 152 (May): 221–36.

Scahill, Rachael I., Chris Frost, Rhian Jenkins, Jennifer L. Whitwell, Martin N. Rossor, and Nick C. Fox. 2003. “A Longitudinal Study of Brain Volume Changes in Normal Aging Using Serial Registered Magnetic Resonance Imaging.” *Archives of Neurology* 60 (7): 989–94.

Schapiro, Anna C., Nicholas B. Turk-Browne, Matthew M. Botvinick, and Kenneth A. Norman. 2017. “Complementary Learning Systems within the Hippocampus: A Neural Network Modelling Approach to Reconciling Episodic Memory with Statistical Learning.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 372 (1711). https://doi.org/10.1098/rstb.2016.0049.

Schapiro, Anna C., Nicholas B. Turk-Browne, Kenneth A. Norman, and Matthew M. Botvinick. 2016. “Statistical Learning of Temporal Community Structure in the Hippocampus.” *Hippocampus* 26 (1): 3–8.

Schlichting, Margaret L., Katharine F. Guarino, Anna C. Schapiro, Nicholas B. Turk-Browne, and Alison R. Preston. 2017. “Hippocampal Structure Predicts Statistical Learning and Associative Inference Abilities during Development.” *Journal of Cognitive Neuroscience* 29 (1): 37–51.

Schlichting, Margaret L., Michael L. Mack, Katharine F. Guarino, and Alison R. Preston. 2019. “Performance of Semi-Automated Hippocampal Subfield Segmentation Methods across Ages in a Pediatric Sample.” *NeuroImage* 191 (May): 49–67.

Schlichting, Margaret L., and Alison R. Preston. 2015. “Memory Integration: Neural Mechanisms and Implications for Behavior.” *Current Opinion in Behavioral Sciences* 1 (February): 1–8.

Schlichting, Margaret L., Dagmar Zeithamova, and Alison R. Preston. 2014. “CA1 Subfield Contributions to Memory Integration and Inference.” *Hippocampus* 24 (10): 1248–60.

Scoville, W. B., and B. Milner. 1957. “Loss of Recent Memory after Bilateral Hippocampal Lesions.” *Journal of Neurology, Neurosurgery, and Psychiatry* 20 (1): 11–21.

Shereen, Ahmed, Niza Nemkul, Dianer Yang, Faisal Adhami, R. Scott Dunn, Missy L. Hazen, Masato Nakafuku, Gang Ning, Diana M. Lindquist, and Chia-Yi Kuan. 2011. “Ex Vivo Diffusion Tensor Imaging and Neuropathological Correlation in a Murine Model of Hypoxia–Ischemia-Induced Thrombotic Stroke.” *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism* 31 (4): 1155–69.

Sierra, Alejandra, Teemu Laitinen, Olli Gröhn, and Asla Pitkänen. 2015. “Diffusion Tensor Imaging of Hippocampal Network Plasticity.” *Brain Structure & Function* 220 (2): 781–801.

Sternberg, R. J., and B. Rifkin. 1979. “The Development of Analogical Reasoning Processes.” *Journal of Experimental Child Psychology* 27 (2): 195–232.

Tamnes, Christian K., Marieke G. N. Bos, Ferdi C. van de Kamp, Sabine Peters, and Eveline A. Crone. 2018. “Longitudinal Development of Hippocampal Subregions from Childhood to Adulthood.” *Developmental Cognitive Neuroscience* 30 (April): 212–22.

Thompson, Carol L., Sayan D. Pathak, Andreas Jeromin, Lydia L. Ng, Cameron R. MacPherson, Marty T. Mortrud, Allison Cusick, et al. 2008. “Genomic Anatomy of the Hippocampus.” *Neuron* 60 (6): 1010–21.

Treit, Sarah, Trevor Steve, Donald W. Gross, and Christian Beaulieu. 2018. “High Resolution in-Vivo Diffusion Imaging of the Human Hippocampus.” *NeuroImage* 182 (February): 479–87.

Venkatesh, Anu, Shauna M. Stark, Craig E. L. Stark, and Ilana J. Bennett. 2020. “Age- and

Memory- Related Differences in Hippocampal Gray Matter Integrity Are Better Captured by NODDI Compared to Single-Tensor Diffusion Imaging." *Neurobiology of Aging* 96 (December): 12–21.

Vodegel Matzen, Mlbl, and Others. 1994. "Performance on Raven's Progressive Matrices. What Makes a Difference?" https://dare.uva.nl/personal/pure/en/publications/performance-on-ravens-progressive-matrices-what-makes-a-difference(dba8a103-851b-4392-b5ca-68082e6730a7).html.

Wechsler, David. 2018. "Wechsler Abbreviated Scale of Intelligence--Second Edition." *PsycTESTS Dataset*. American Psychological Association (APA). https://doi.org/10.1037/t15171-000.

Wendelken, Carter, and Silvia A. Bunge. 2010. "Transitive Inference: Distinct Contributions of Rostrolateral Prefrontal Cortex and the Hippocampus." *Journal of Cognitive Neuroscience* 22 (5): 837–47.

Wimmer, G. Elliott, and Daphna Shohamy. 2012. "Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions." *Science* 338 (6104): 270–73.

Wolf, Dominik, Florian U. Fischer, Robin de Flores, Gaël Chételat, and Andreas Fellgiebel. 2015. "Differential Associations of Age with Volume and Microstructure of Hippocampal Subfields in Healthy Older Adults." *Human Brain Mapping* 36 (10): 3819–31.

Zeithamova, Dagmar, April L. Dominick, and Alison R. Preston. 2012. "Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference." *Neuron* 75 (1): 168–79.