Supplement

Methods

Table S1

Detailed Trial Order for Both Sessions

Object-Location	Trial Order												
	Learning Session												
1-Bike	T	T	T	P	T	EP	T	AP	T	EP	T	AP	T
2-Soccer Ball	T	T	T	P	T	EP	T	AP	T	EP	T	AP	T
3-Violin	T	T	T	P	T	EP	T	AP	T	EP	T	AP	T
1-Bike	EP	AP											
3-Violin	EP	AP											
2-Soccer Ball	EP	AP											
	Delayed Session												
2-Soccer Ball	EP	AP	EP	AP	P								
3-Violin	EP	AP	EP	AP	P								
1-Bike	EP	AP	EP	AP	P								

Note. T = training trial; P = probe trial (with same configuration as in T); EP = egocentric probe trials (same start, landmarks invisible); AP = allocentric probe trial (new start, landmarks visible); the object-location pairing (e.g., bike at location 1) was randomized and the order of presentation (e.g., 1-bike pair learned in first block) was pseudorandomized; n = 45 trials in learning session; n = 15 trials in delayed session.

A: Data Pre-processing

Navigation data

1. *Memory score* [%] (Bellmund et al., 2020; Iggena et al., 2023). For each trial, we calculated the memory error as the Euclidian distance between goal location and remembered location. We compared this value to a reference distribution, which was obtained by computing the Euclidian

distance between each goal location and 1000 randomly sampled locations. The memory score corresponds to the percentile rank on the error distribution i.e., the proportion of random locations that were further away from the goal location than the remembered goal location. Thus, the memory score ranges from zero to one, with one indicating high memory performance and 0.5 corresponding to chance level. A value below 0.5 is possible and indicates a systematic bias in the opposite direction.

5. *Initial rotation* [rad] (also called integrated absolute angular velocity (IdPhi) or vicarious-trial-and-error behavior (VTE) (Iggena et al., 2023; Santos-Pata & Verschure, 2018; Schmidt et al., 2013). Here, we used the z-rotation data from the beginning of the trial up until when the participant left the first decision point (intersection). The z-rotation raw data was transformed from degrees to radians, unwrapped to avoid circular transitions when calculating change and summed up. Some researchers divide this value by time (velocity), which we decided against because children tended to take more time to start and complete the trials than adults, potentially biasing velocity measures. Higher initial rotation values denote more visual exploration at the beginning of the trial.

Spatial knowledge data

For evaluating the landmark position task, we used the Gardony Map Drawing Analyzer software (GMDA; Gardony et al., 2016). We took a screenshot of the participants map arrangement. First, we manually marked the positions of all correctly identified environmental landmarks and goal objects in GMDA. The authors of GMDA recommend rotating the participant’s map in GMDA in small steps to determine the orientation with a minimal theta value. Based on the ideal orientation, the software automatically performs a pairwise comparison of objects in the participant’s map with the template map. We obtained three scores: First, canonical accuracy, second, distance accuracy, reflecting the magnitude of inter-landmark distance error and third, angle accuracy, reflecting the magnitude of inter-landmark angular error. The three scores were averaged into a single composite score, ranging from zero to one with larger values denoting higher overall accuracy.

B: Statistical analysis

Model covariate: Joystick control abilities

There were significant age group differences in the ability to use a joystick during practice. Specifically, 6-8YO children required more time to solve the task ($F(2,93) = 21.16$, $p < .001$; 6-8YO vs. 9-11YO: $t(93) = 4.56$, $p_{\mathrm{Bonferroni(3)}} < .001$; 6-8YO vs. AD: $t(93) = 6.28$, $p_{\mathrm{Bonferroni(3)}} < .001$) and had longer excess paths ($F(2,93) = 22.04$, $p < .001$; 6-8YO vs. 9-11YO: $t(93) = 4.94$, $p_{\mathrm{Bonferroni(3)}} < .001$; 6-8YO vs. AD: $t(93) = 6.28$, $p_{\mathrm{Bonferroni(3)}} < .001$). We wanted to account for this potential confounder and computed a composite score by averaging the z-scored latency and z-scored excess path length during practice that we used as a model covariate in all our LMM.

Random effects in Linear Mixed Models

We modelled all LMM with the maximal random effects structure (Barr, 2013) i.e., including a random intercept, random slopes for all within-subjects factors and correlations among the random effects. In case of non-convergence, we simplified the random effects structure, meaning that we performed a stepwise removal starting with the highest-order random effects, until the model converged (Matuschek et al., 2017). To determine if the inclusion of a random slope for session or condition significantly improved the model fit, we used Likelihood-Ratio-Tests (LRT).

Results

Table S2

Descriptive Values (Mean and SD) for All Variables

Variable	6-8YO		9-11YO		AD	
	M	SD	M	SD	M	SD
			Learning Session – Training Trials			
Latency	22.51	15.17	16.24	11.27	13.06	6.77
Excess Path Length	42.03	59.76	22.82	43.42	10.70	19.33
Excess Distance to Goal	4.21	6.02	2.90	4.70	2.45	3.31
Initial rotation	0.93	0.58	0.86	0.45	0.86	0.99
			Learning Session – Probe Trials			
Egocentric Memory Score	0.74	0.28	0.90	0.19	0.98	0.07
Allocentric Memory Score	0.74	0.30	0.83	0.28	0.96	0.13
			Delayed Session – Probe Trials			
Egocentric Memory Score	0.60	0.31	0.71	0.30	0.89	0.21
Allocentric Memory Score	0.59	0.32	0.68	0.33	0.85	0.25
			Delayed Session – Spatial Knowledge			
Boundary Score	0.13	0.34	0.47	0.51	0.81	0.40
Landmark Identity Score	0.65	0.21	0.74	0.19	0.70	0.25
Landmark Position Score	0.73	0.07	0.72	0.07	0.82	0.12

Note. 6-8YO = 6-8-year-old children; 9-11YO = 9-11-year-old children; AD = adults. M = mean; SD = standard deviation.

A: Above-chance level performance for immediate spatial memory in the learning session

Participants in all age groups reached above chance-level performance for the memory score in the learning session. Tests against 0.5 for 6-8YO children: $\hat{\theta} = 0.76$, $95\%\ \mathrm{CI}_{\mathrm{Bonferroni(3)}}\ [0.71, 0.80]$, $t(94.17) = 13.90$, $p_{\mathrm{Bonferroni(3)}} < .001$, 9-11YO children: $\hat{\theta} = 0.86$, $95\%\ \mathrm{CI}_{\mathrm{Bonferroni(3)}}$ $[0.82, 0.90]$, $t(90.53) = 21.20$, $p_{\mathrm{Bonferroni(3)}} < .001$, and adults: $\hat{\theta} = 0.96$, $95\%\ \mathrm{CI}_{\mathrm{Bonferroni(3)}}$ $[0.91, 1.00]$, $t(91.70) = 25.36$, $p_{\mathrm{Bonferroni(3)}} < .001$.

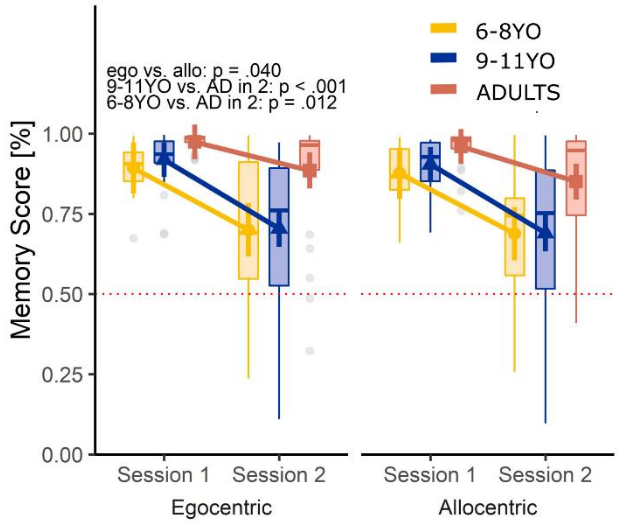
B: Spatial memory consolidation after two weeks for well-learned object-location pairs

Children show worse navigation-based memory consolidation than adults for well-learned pairs

We ran a control analysis to further explore if encoding differences account for the absence of age-related differences in spatial memory consolidation. Here, we computed a LMM with the memory score, including only object-location pairs that were well-learned in the first session. Well-learned was defined as finding the correct goal (i.e., pressing the button in the correct alley or the adjacent intersection) in at least two out of three learning session probe trials in both egocentric and allocentric conditions. Trials with object-location pairs that did not meet this criterion were excluded. This reduced our sample to n = 22 6-8YO, n = 29 9-11YO, and n = 31 adults. The results paralleled our previous analysis, namely we found main effects for age group ($p = .001$; see statistics in Table S2, Figure S1), condition ($p = .040$), and session ($p < .001$). In addition, we did find evidence for an interaction between age group and session for well-learned pairs ($p = .028$). Post-hoc contrasts showed that while there were no age-related performance differences for well-learned pairs during learning, adults remembered them significantly better than both 6-8YO children ($t(115.92) = -3.29$, $p_{\mathrm{Bonferroni(9)}} = .012$) and 9-11YO children ($t(131.14) = -4.49$, $p_{\mathrm{Bonferroni(9)}} < .001$) after the two-week delay period. This implies that adults profit from superior consolidation compared to children when having a comparable level of encoding. The LMM had a significantly better fit when including the random slope for session ($\chi^2(1) = 190.22$, $p < .001$), meaning that there was interindividual variation in participant's forgetting rates across the two-week delay period.

Figure S1

Spatial Memory Consolidation of Well-Learned Pairs after Two Weeks



Note. Significant advantage for egocentric over allocentric memory across age groups and sessions. While there were no age-related performance differences for well-learned pairs during the first session, adults remembered them significantly better than both 6-8YO in the second session after the two-week delay period (across conditions).

Table S3

Fixed and Random Effects of LMM for Probe Trials across Sessions (Well-Learned Pairs)

Variable	Memory Score (Well-Learned Pairs)				
	Fixed Effects				
	F	df^{S}	$df^{\mathrm{S}}_{\mathrm{res}}$	p	ω^2
Group	7.37	2	73.54	.001	0.14
Session	74.17	1	81.97	< .001	0.47
Condition	4.20	1	1,780.98	.040	0.00
Covariate Sex	0.01	1	71.59	.940	-0.01
Covariate Motor Score	0.00	1	76.23	.945	-0.01
Group × Session	3.75	2	80.98	.028	0.06
Group × Condition	0.20	2	1,780.94	.816	0.00
Session × Condition	0.09	1	1,780.99	.769	0.00
Group × Session × Condition	0.23	2	1,780.95	.792	0.00
	Random Effects				
	SD	r	p		
Participant ID (intercept)	0.10				
Session (slope)	0.08		< .001		
Participant ID × Session					
Residual	0.18				