

## Supplementary Materials

### Comparison of ASD and TD groups

We compared the performance of ASD children in our sample to that of  $N = 34$  6- to 11-year-old children drawn from Study 2 of Ruggeri et al. (2019; note that the original sample included 4- and 5-year-olds who are excluded from the present analysis to better match the groups). These participants were tested using a procedure identical to that described in the present study. Descriptive statistics for performance in the item recognition and visuospatial memory tests are displayed in Table S1, as well as within-subject differences between active and yoked study.

To compare overall performance between groups we conducted a signal detection analysis for both item recognition and visuospatial memory, resulting in discriminability ( $d'$ ) and bias ( $c$ ) scores (Table S2). Mixed effects models were fit for each measure with group (TD/ASD), test (test/retest), and the group X test interaction as fixed effects, with random intercepts for participants. Analysis of deviance was used to identify significant effects within each model.

**Item recognition.** There was a main effect of group on discriminability,  $\chi^2(1) = 7.30, p = .007$ , such that discriminability was higher among TD children. Overall, discriminability decreased from test to retest,  $\chi^2(1) = 12.27, p < .001$ , but there was no group X test interaction,  $\chi^2(1) = 1.33, p = .25$ . The corresponding analysis on bias ( $c$ ) scores indicated a marginally significant shift toward conservative responding from test to retest,  $\chi^2(1) = 2.85, p = .09$ , but no effect of group,  $\chi^2(1) = .62, p = .43$ , and no group X test interaction,  $\chi^2(1) = 2.39, p = .12$ .

In order to directly compare the effects of active study across groups, a new mixed effects logistic regression model of recognition accuracy for studied items was fit with group (TD/ASD), condition (active/yoked), false alarm rate, and the group X condition interaction as fixed effects. This analysis confirmed that among the TD group recognition accuracy was significantly lower

following yoked study (OR = 0.47 [.41, .55]) compared to active study. In addition, the disadvantage from yoked study was significantly larger in the TD group than in the ASD group (OR = 0.69 [.56, .85]).

**Visuospatial memory.** There were no differences between active and yoked study for spatial location memory in either the present study or Study 2 of Ruggeri et al. (2019), so our analysis focused on comparing overall spatial memory between the ASD and TD groups. A signal detection analysis was performed on participants' ability to judge whether a studied item appeared in the same or different location as compared to the study phase. Hits were defined as correctly identifying that an item appeared in the same location. In contrast to item recognition, discriminability ( $d'$ ) was relatively low for location memory in both TD and ASD children, particularly in the delayed retest (Table S2). There was a significant main effect of testing session,  $\chi^2(1) = 21.38, p < .001$ , such that discriminability decreased from test to retest. There was no effect of group,  $\chi^2(1) = 1.25, p = .26$ , and no group X test interaction,  $\chi^2(1) = 1.91, p = .17$ .

The corresponding analysis for response bias revealed a main effect of group,  $\chi^2(1) = 8.82, p = .003$ , such that TD participants were more liberal (i.e., they were more likely to respond that an object appeared in the same location as during study). There was no effect of testing session,  $\chi^2(1) = .19, p = .66$ , and no group X test interaction,  $\chi^2(1) = .08, p = .77$ .

Table S1. Descriptive statistics for item and spatial location performance in TD comparison group.

	Active	Yoked	Active - Yoked
<i>Item recognition (hit rates)</i>			
Test	.77 (.20)	.64 (.19)	.13 (.17)
Retest	.69 (.23)	.56 (.24)	.13 (.17)
<i>Spatial location (proportion correct)</i>			
Test - Same location	.90 (.12)	.91 (.14)	-.004 (.12)
Test - Different location	.24 (.32)	.27 (.28)	-.02 (.16)
Retest - Same location	.84 (.21)	.84 (.21)	-.007 (.12)
Retest - Different location	.18 (.23)	.19 (.26)	-.01 (.18)

Table S2. Descriptive statistics from signal detection analysis of ASD and TD groups.

	Item recognition (old/new)		Spatial location (same/different)	
	Sensitivity (d')	Bias (c)	Sensitivity (d')	Bias (c)
<i>Test</i>				
TD	2.51 (.86)	0.67 (.52)	0.49 (.81)	-1.12 (.69)
ASD	1.89 (.89)	0.70 (.53)	0.76 (.61)	-0.60 (.74)
<i>Retest</i>				
TD	2.14 (.66)	0.71 (.63)	0.13 (.37)	-1.12 (.78)
ASD	1.72 (.89)	0.84 (.40)	0.09 (.54)	-0.67 (.82)