Supplemental Results

Performance

Performance on the test trials in each condition of the CFMT (immediate or delayed recognition) was examined separately with 2 (group) x 3 (age group) univariate ANOVAs, including post-hoc Tukey's test for age. Based on our *a priori* hypotheses, follow-up analyses included one-way ANOVAs for age performed separately in each group with post-hoc Tukey's test for age, and independent-samples t-tests for group differences performed separately in each age group. We also analyzed the performance data using an arcsine transformation to correct for deviations from normality, but this transformation did not affect the results. Therefore, for clarity, the results are presented here without the transformation.

Immediate recognition. Performance on the immediate recognition condition was examined with a univariate ANOVA with percent correct as the dependent variable and group (TD, autism) and age (children, adolescents, adults) as fixed factors. The ANOVA revealed a significant main effect of group ($F(1, 140)= 22.442$, $p< 0.001$, $\eta^2_p= 0.138$), with the TD group performing more accurately than the group with autism. There was no significant main effect of age ($p= 0.156$) of interaction of group and age ($p= 0.101$).

One-way ANOVAs performed separately for each group, with percent correct as the dependent variable and age as a fixed factor, indicated that performance improved significantly with age in the TD group ($F(2, 79)= 9.898$, $p< 0.001$, $f= 0.507$), driven by a significant increase

in accuracy from childhood to adulthood ($p < 0.001$). Age-related changes did not occur in the group with autism ($p = 0.425$). Independent-samples t-tests performed separately for each age revealed a significant between-group difference in adulthood, with TD adults performing better than adults with autism ($t(18.270) = 3.457$, $p = 0.002$, $d = 1.149$; Figure A1a). Performance did not significantly differ between groups in childhood or adolescence (though there was a trend toward better performance in the TD groups, p's > 0.06), an unexpected result also evident in O'Hearn et al., (2010; see also Croydon et al., 2014). There was no main effect of age or interaction of group and age (p's > 0.101), but this may reflect ceiling performance typically.

Delayed recognition. A similar univariate ANOVA examined performance on the delayed recognition condition. The ANOVA revealed a significant main effect of group ($F(1, 140) = 23.960$, $p < 0.001$, $\eta^2_p = 0.146$), with the TD group performing more accurately than the group with autism. A main effect of age was also present ($F(2, 140) = 10.095$, $p < 0.001$, $\eta^2_p = 0.126$). Post-hoc Tukey's tests indicated that, across groups, performance improved from childhood to adolescence ($p = 0.064$), from adolescence to adulthood ($p = 0.022$), and overall from childhood to adulthood ($p < 0.001$). A significant group x age interaction ($F(2, 140) = 3.457$, $p = 0.034$, $\eta^2_p = 0.047$) indicated that group differences changed with age (Figure A1b).

One-way ANOVAs performed separately for each group, with percent correct as the dependent variable and age as a fixed factor revealed that performance on the delayed recognition condition improved with age in both groups (TD, $F(2, 79) = 10.007$, $p < 0.001$, $f = 0.510$; autism, $F(2, 65) = 3.906$, $p = 0.025$, $f = 0.352$). In the TD group, performance improved

from childhood to adulthood (*p*= 0.001), and from adolescence to adulthood (*p*= 0.010), reflecting the typical late development in face recognition ability. In the group with autism, however, performance improved from childhood to adolescence (*p*= 0.028), but not from adolescence to adulthood (*p*= 0.901) or overall from childhood to adulthood (*p*= 0.107). This indicates that face recognition is improving in those with autism in late childhood, but these gains are actually lost by adulthood. Independent-samples t-tests performed separately for each age revealed group differences in childhood (*t*(41.571)= 2.894, *p*= 0.006, *d*= 0.775) and adulthood (*t*(45)= 4.303, *p*< 0.001, *d*= 1.289), with the TD groups performing significantly better than their peers with autism; group differences were not present in adolescence (*p*= 0.223).
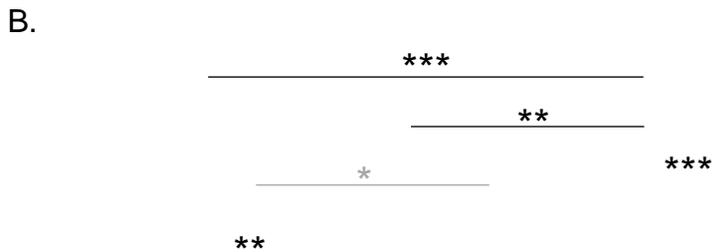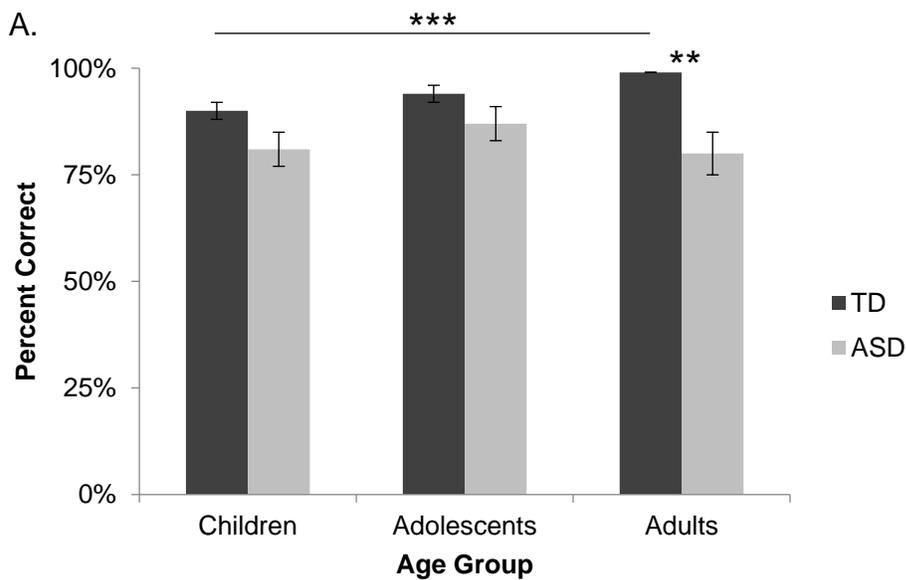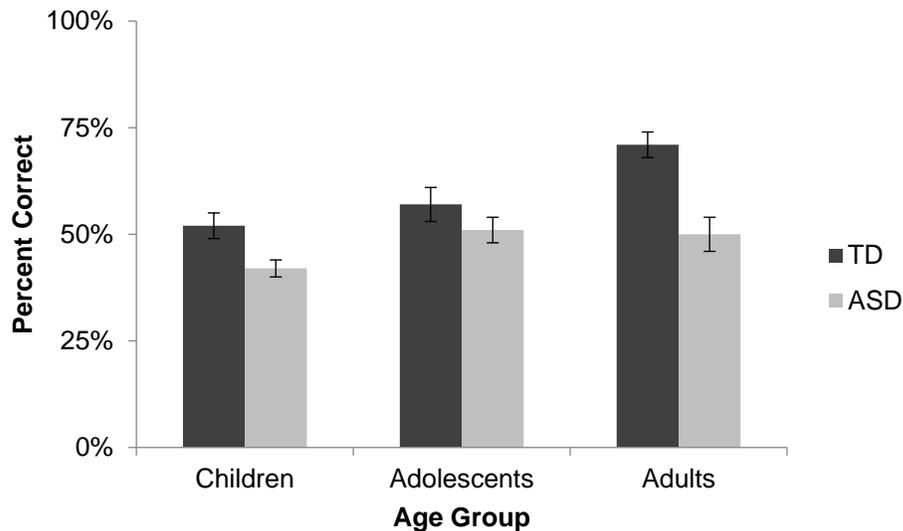
Figure A1. (a) Performance on immediate recognition test trials improved typically (TD; dark

grey bars) from childhood to adulthood, but did not change with age in autism (ASD; light grey

bars), resulting in significant group differences by adulthood. (b) Performance on delayed

recognition test trials improved with age in both groups, however, this improvement occurred

early in development in autism, but continued into adulthood typically. TD children and adults

performed significantly better on the task than did their peers with autism, though groups did not

differ in adolescence

Summary

Consistent with previous behavioral work (O'Hearn et al., 2010), we found that

performance on the CFMT improved with age typically but not in the group with autism.

Performance on the immediate recognition condition improved from childhood to adulthood typically, but not in autism, resulting in a significant disparity in face recognition ability in adulthood. Performance on the delayed recognition condition improved from childhood to adolescence in *both* groups. However, performance continued to improve from adolescence to adulthood in the TD group, but did not in those with autism. Between-group differences were present in childhood and adulthood, with the TD group performing more accurately than their peers with autism, although groups did not differ in adolescence. These results are consistent with the comprehensive review by Weigelt et al. (2012), who found that, although simple face perception is relatively unaffected in autism, face memory *is* impacted in autism, and poorer performance is evident even on tasks that impose only a minimal demand on memory.

Unlike previous work with the CFMT, we found group differences present in childhood in our current sample of participants. The average IQ of the current sample was slightly higher than in previous work, which may have minimized floor effects and/or decreased variation in younger participants in the present study. In addition, the current sample was larger, increasing our statistical power. This finding is consistent, however, with more recent work using a version of the CFMT designed specifically for children (CFMT-C; Croydon et al., 2014). Children aged 7-12 with autism scored significantly lower on the CFMT-C than did a group of TD children of similar age and IQ. Additionally, other work that has investigated face recognition in autism using easier tasks has also shown group differences in childhood and adolescence, although difficulties increase into adulthood (O'Hearn et al., 2014).

Eye movement measures: remainder of face and nose ROIs

We examined patterns of fixation to the remainder of the face (with other ROIs subtracted; i.e., non-featural areas including the cheeks, forehead and chin) and the nose. We first performed univariate ANOVAs with either average fixation duration or fixation count as the dependent variable with group (TD, autism) and age (children, adolescents, adults) as fixed factors. We performed the ANOVAs separately for each ROI (remainder of face, nose) within each condition (immediate recognition, delayed recognition) and trial type (memory, test). Follow-up analyses included one-way ANOVAs performed separately in each group to examine age-related changes in fixations to the face and nose typically and in autism, and independent-samples t-tests performed separately in children, adolescents, and adults to examine between-group differences in patterns of fixation to these ROIs at each age.

Patterns of fixation to remainder of face

Immediate recognition: memory trials

Univariate ANOVAs did not reveal any significant effects or interaction of group or age on either average fixation duration or fixation count for the remainder of the face during immediate recognition memory trials (p's > 0.430).

Immediate recognition: test trials

Average fixation duration. A univariate ANOVA with average fixation duration for the remainder of the face during immediate recognition test trials as the dependent variable and group (TD, autism) and age (children, adolescents, adults) as fixed factors revealed a significant group x age interaction (F(2, 140)= 3.096, p= 0.048, η^2_p= 0.042), although main effects of group and age did not approach significance (p's > 0.253). Follow-up one-way ANOVAs examining age-related changes in average fixation duration for the remainder of the face separately in each group revealed a marginally significant main effect of age in the TD group (F(2, 79)= 2.798, p= 0.067, f= 0.270), but not in the group with autism (p= 0.184). Post-hoc Tukey's tests revealed that the effect in the TD group was driven by trend-level decreases in average fixation duration for the remainder of the face typically from adolescence to adulthood (p= 0.055). Follow-up independent samples t-tests indicated a significant between-group difference in adulthood (t(23.782)= -2.766, p= 0.011, d= 0.869), with longer average fixation duration for the remainder of the face in adults with autism as compared to TD adults. Groups did not significantly differ in childhood or adolescence (p's > 0.207).

Fixation count. A similar univariate ANOVA with fixation count for the remainder of the face during immediate recognition test trials revealed a significant main effect of group (F(1, 140)= 5.934, p= 0.016, η^2_p= 0.041), with the group with autism making more fixations to the remainder of the face than the TD group. There was no significant main effect of age or interaction of group and age (p's > 0.525). Follow-up independent-samples t-tests indicated

between-group differences at trend level in childhood (t(51)= -1.866, p= 0.068, d= 0.512), and

adolescence (t(44)= -1.751, p= 0.087, d= 0.516), with higher fixation count for the remainder of

the face in the groups with autism relative to their TD peers; groups did not differ in adulthood,

p= 0.768.

Delayed recognition: memory trial

Average fixation duration. A univariate ANOVA with average fixation duration for the

remainder of the face during the delayed recognition memory trial as the dependent variable did

not reveal any effects or interaction of group or age (p's > 0.189).

Fixation count. A univariate ANOVA with fixation count for the remainder of the face

during the delayed recognition memory trial as the dependent variable revealed a main effect of

age (F(2, 140)= 2.652, p= 0.074, η^2_p= 0.037). Post-hoc Tukey's tests revealed that this effect was

driven by a marginally significant decrease in fixation count for the remainder of the face from

childhood to adolescence across groups (p= 0.059). However, follow-up one-way ANOVAs did

not reveal any age-related changes in fixation count for the remainder of the face in either the TD

group or the group with autism (p's > 0.139). There was no significant main effect of group or

interaction of group and age (p's > 0.467).

Delayed recognition: test trials

Univariate ANOVAs did not reveal any significant effects or interaction of group or age

on either average fixation duration or fixation count for the remainder of the face during delayed recognition test trials (p's > 0.126).

Patterns of fixation to noses

Immediate recognition: memory trials

Univariate ANOVAs did not reveal any significant effects or interaction of group or age on either average fixation duration or fixation count for the nose during immediate recognition memory trials (p's > 0.396).

Immediate recognition: test trials

Univariate ANOVAs did not reveal any significant effects or interaction of group or age on either average fixation duration or fixation count for the nose during immediate recognition test trials (p's > 0.216).

Delayed recognition: memory trial

Average fixation duration. A univariate ANOVA with average fixation duration for the nose during the delayed recognition memory trial as the dependent variable and group (TD, autism) and age (children, adolescents, adults) as fixed factors revealed a trend-level main effect

of group (F(1, 89)= 3.426, p= 0.067, η^2_p= 0.037), with the TD group making longer fixations to the nose as compared to the group with autism. Follow-up independent-samples t-tests indicated a between-group difference in adolescence (t(18.948)= 2.431, p= 0.033, d= 0.852), again with longer average fixation duration for the nose in the TD group as compared to the group with autism; groups did not differ in childhood or adulthood (p's > 0.281). The initial ANOVA did not reveal a significant main effect of age or interaction of group and age (p's > 0.421).

Fixation count. A similar univariate ANOVA with fixation count for the nose during the delayed recognition memory trial as the dependent variable did not reveal any significant effects or interaction of group or age (p's > 0.378).

Delayed recognition: test trials

Univariate ANOVAs did not reveal any significant effects or interaction of group or age on either average fixation duration or fixation count for the nose during delayed recognition test trials (p's > 0.226).

Summary

Patterns of fixations to the remainder of the face did not differ between groups or age groups during the immediate recognition memory trials or the delayed recognition test trials. During the immediate recognition test trials, average fixation duration for the remainder of the

face changed with age in the TD group, decreasing from adolescence to adulthood, but did not

change with age in the group with autism. This resulted in a significant group difference in

adulthood, with adults with autism making longer fixations to these non-featural areas as

compared to their TD peers. This is consistent with Pelphrey et al. (2002), who found that adults

with autism spent a greater percentage of time fixating areas outside the core face features as

compared to TD adults. Additionally, group differences were present for fixation count during

the immediate recognition trials in childhood and adolescence, with the groups with autism

making more fixations to the remainder of the face as compared to the TD groups. During the

delayed recognition memory trial, fixation count for the remainder of the face changed with age

across both groups, decreasing from childhood to adolescence.

Patterns of fixation for the nose did not differ between groups or change with age during

the majority of the CFMT. However, during the delayed recognition memory trial, average

fixation duration for the nose was longer in the TD group as compared to the group with autism.

This trend was driven by a group difference in adolescence, with TD adolescents making longer

fixations to the nose than adolescents with autism. Group differences for the remainder of face

and nose ROIs are also summarized in Table A1.

Table A1

Summary of group differences for remainder of face and nose in each condition and trial type

Condition	Trial Type	Remainder of Face				Nose			
		Average Fixation Duration	p	Fixation Count	p	Average Fixation Duration	p	Fixation Count	p
Immediate Recognition	Memory								
	Test	Adult: ASD > TD	.01	Child: ASD > TD Adol: ASD > TD	.07 .09				
Delayed Recognition	Memory					Adol: TD > ASD	.03		
	Test								