Supplementary Information

Table S1 Inter-trial Interval and Valid Trial Number of Experiment 1 & 2

Experiment		Autistic M±SD	NT M±SD	*t*	*P*
Experiment 1	Inter-trial interval (s)	15.57±5.47	13.94±2.96	1.73	0.09
	Valid trial number				
	Mild sound	8.49±1.17	9.39±0.98	–3.84	<0.001
	Aversive sound	8.46±1.12	9.13±1.19	–2.63	0.01
Experiment 2	Inter-trial interval (s)	14.42±5.60	13.48±3.15	0.92	0.36
	Valid trial number				
	Non-startle sound	8.34±1.41	9.15±1.17	–2.71	0.008
	Startle sound	8.37±1.21	9.32±0.93	–3.84	<0.001

Pupil Data Preprocessing

Data analysis was conducted in MATLAB and R. First, the pupil diameters of the left and the right eyes were averaged only when both eyes were tracked. Data points where only one eye was tracked were discarded and labeled as missing data. Second, the pupil signal was interpolated according to Nuske et al. (2014). To be brief, samples with change velocity larger than two standard deviations of the mean, calculated by each participant, were removed to reduce sharp spikes due to blinks and partial eyelid closures. A gap of missing data was recovered with linear interpolation if it was less than 350 ms and between stable traces, which contained at least 50% tracked sample around the gap (see Nuske et al., 2014 for details). After pupil interpolation, the signal was smoothed using a Savitzky-Golay filter. We excluded invalid trials that contained less than 60% valid pupil data during the sound presentation or participants with less than six valid trials in any of the conditions.

Trial-by-Trial Analysis of Pupil Data Validity

In each trial, we calculated pupil data validity (in all trials including valid and invalid ones) when sound present as a measurement of data quality. In each experiment, we applied a generalized estimation equation (GEE) model (Halekoh et al., 2006) on pupil data validity with the main effect of group, trial, and the interaction between them as predictors. As shown in Fig. S1, in both Experiment 1 and 2 the overall pupil data quality was lower in the autistic group than the NT group as expected, *P*s<0.001. However, the main effects of trial and the interaction effects between group and trial were not significant, *P*s>0.29. That is, in both experiments we did not observe a systematic change of pupil data quality over trials in both groups.

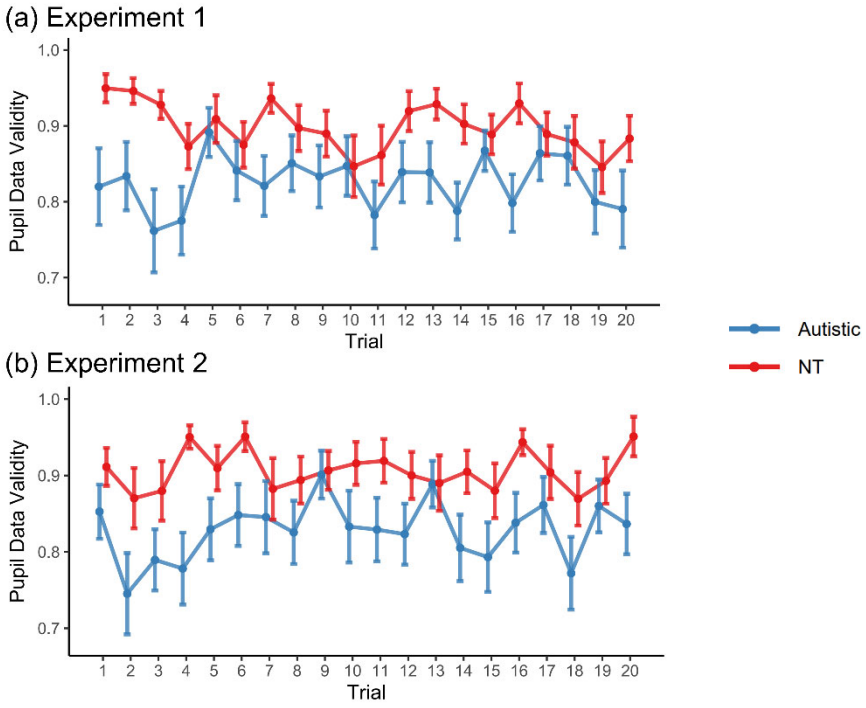

Fig. S1 Pupil data validity over trials in (a) Experiment 1 and (b) Experiment 2
In both experiments, the pupil data validity was lower in the autistic group than the NT group and did not systematically change over trials in both groups

Trial-by-Trial Analysis of Baseline Pupil Size

In each experiment, we applied a GEE model (Halekoh et al., 2006) on the 1 s averaged baseline pupil diameter before sound onset with the main effect of group, trial, and the interaction between them as predictors. As shown in Fig. S2, in Experiment 1 and 2 we found that the baseline pupil diameter increased over the course of the experiment, *P*s<0.02. However, the interaction effect between group and trial as well as the main effect of group were not significant in both experiments, *P*s>0.37. That is, the baseline pupil diameter increased over trials with similar rates and stayed comparable between the autistic group and the NT group.

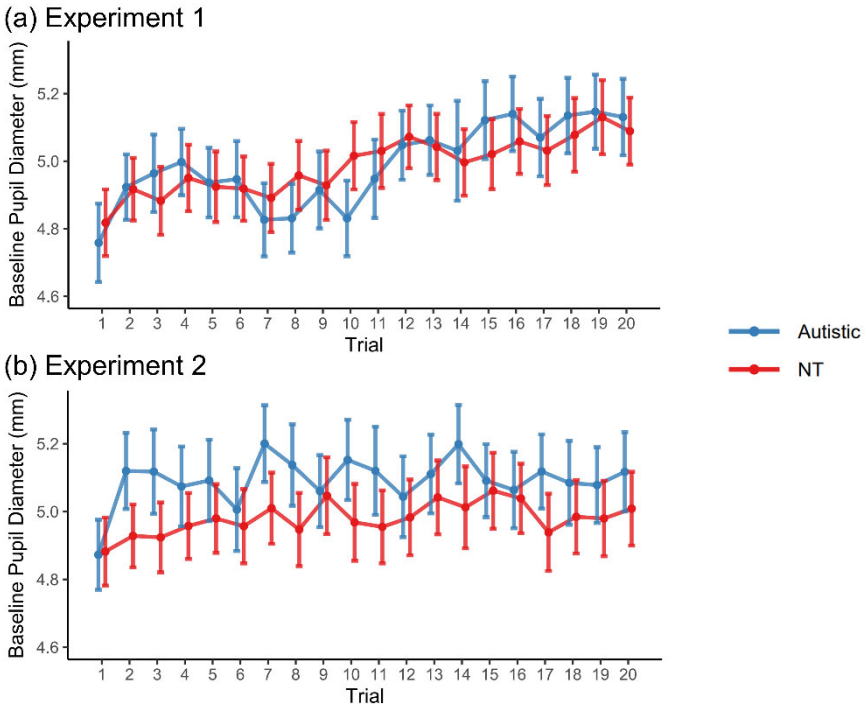

Fig. S2 Baseline pupil diameter over trials in (a) Experiment 1 and (b) Experiment 2
In both experiments, the baseline pupil diameter increased over trials with similar rates and stayed comparable between the autistic group and the NT group

References

Halekoh U, Højsgaard S, Yan J, 2006. The R package geepack for generalized estimating equations. *J Stat Soft*, 15(2):1–11. https://doi.org/10.18637/jss.v015.i02

Nuske HJ, Vivanti G, Dissanayake C, 2014. Reactivity to fearful expressions of familiar and unfamiliar people in children with autism: an eye-tracking pupillometry study. *J Neurodev Disord*, 6(1):14. https://doi.org/10.1186/1866-1955-6-14