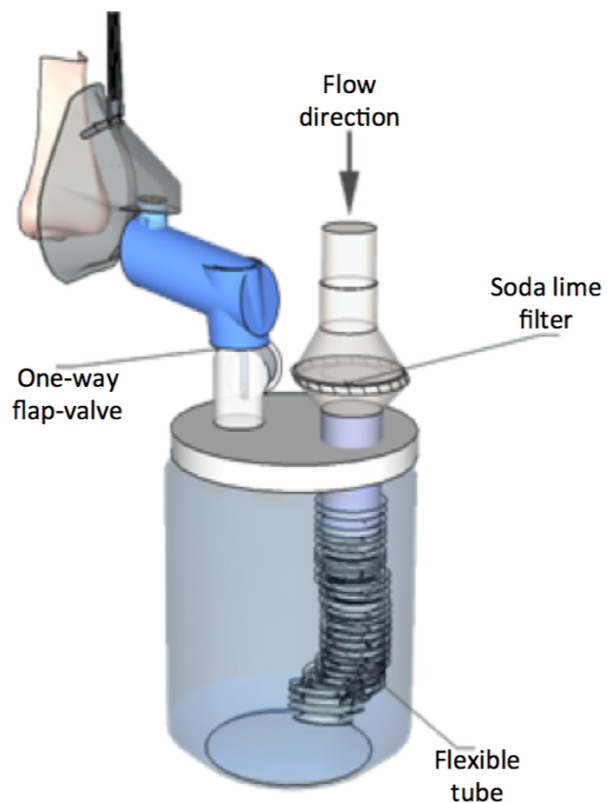


In the format provided by the authors and unedited.

Altered responses to social chemosignals in autism spectrum disorder

Yaara Endevelt-Shapira ^{*}, Ofer Perl, Aharon Ravia , Daniel Amir, Ami Eisen, Vered Bezalel, Liron Rozenkrantz, Eva Mishor, Liron Pinchover, Timna Soroka, Danielle Honigstein and Noam Sobel ^{*}

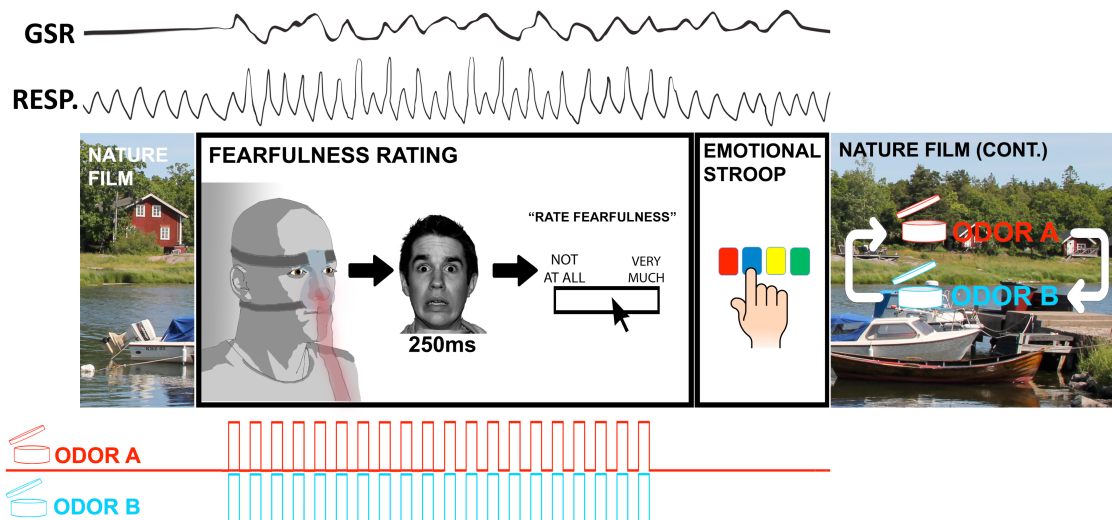
Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel. *e-mail: yaara.yaara@gmail.com; noam.sobel@weizmann.ac.il



Supplementary Figure 1

Illustration of the shirt-sniffing device (SSD)

To standardize body-odor sampling we developed the SSD. This consisted of a glass jar containing the T-shirt, with an air intake port via soda lime filter, and air sampling port via one-way flap valve into individual-use airtight nose mask. This arrangement assured that environmental odors didn't contaminate the sample during the sampling process. The recognizable person in the figure is a co-author and not a participant.

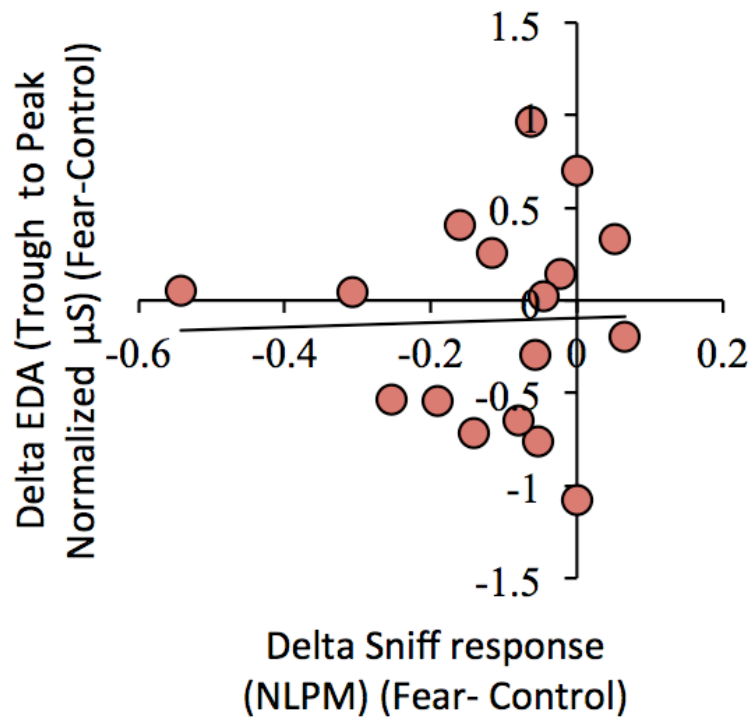


Supplementary Figure 2

Experimental time-course for Stroop and Faces experiments with fear sweat

Each participant tested twice on the same day, once with the fear sweat and once with control, counterbalanced for order, participants blind to condition. Fearful face image from the NimStim set of facial expressions¹. All other photos and graphics: Ofer Perl.

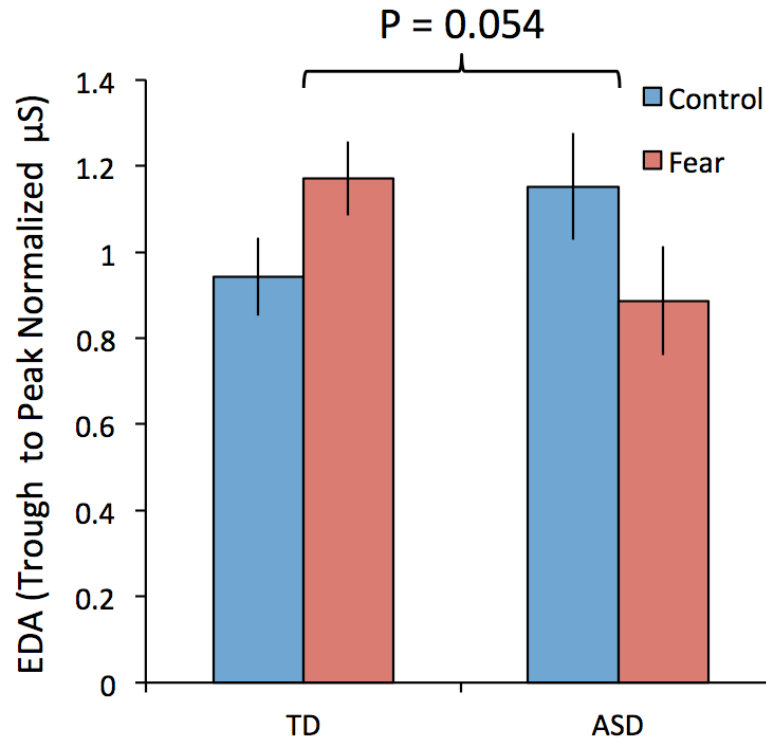
1. Tottenham, N., Tanaka, J., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., Nelson, C.A. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, 168(3):242-9.



Supplementary Figure 3

ER-EDA effects not driven by sniffing

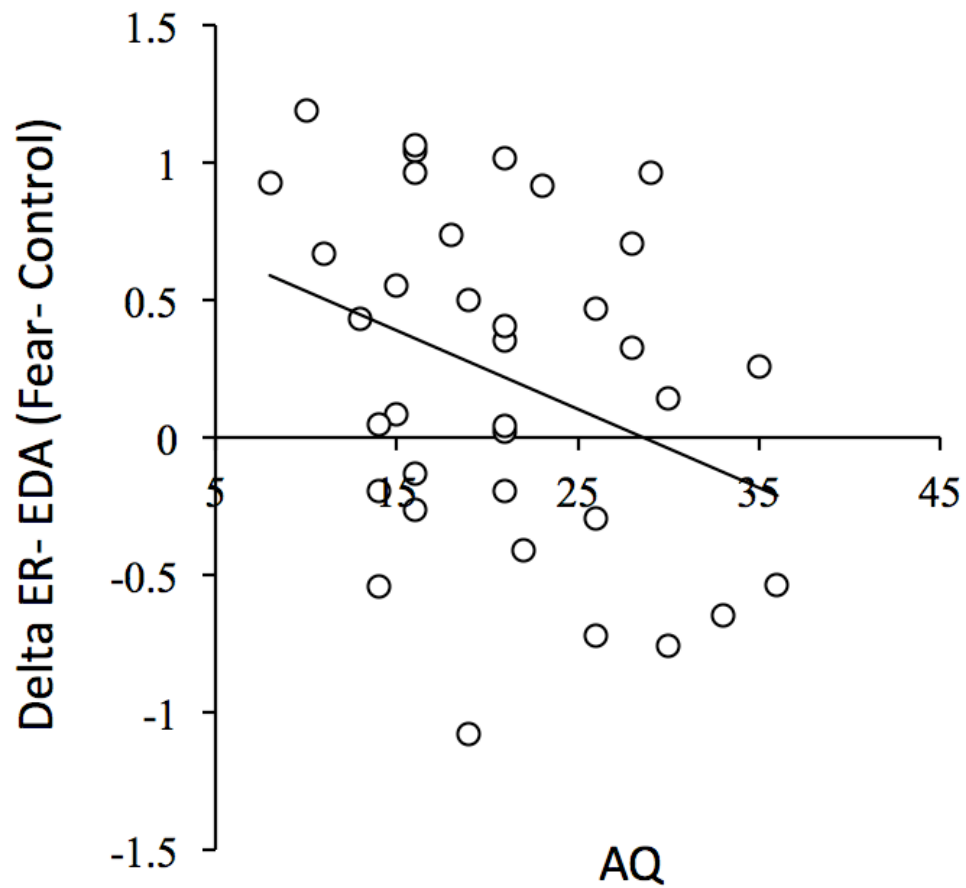
The extent of change in ER-EDA as a function of the extent of change in sniff airflow in ASD. There was a slight trend for increased sniff modulation to the smell of fear in ASD vs. TD. Thus, one could raise the possibility that this is driving the ER-EDA effects. To address this possibility, we tested for a relation between change in ER-EDA and change in nasal airflow in ASD, yet observed no relation ($r = 0.033$, $p = 0.9$).



Supplementary Figure 4

ER-EDA during Stroop

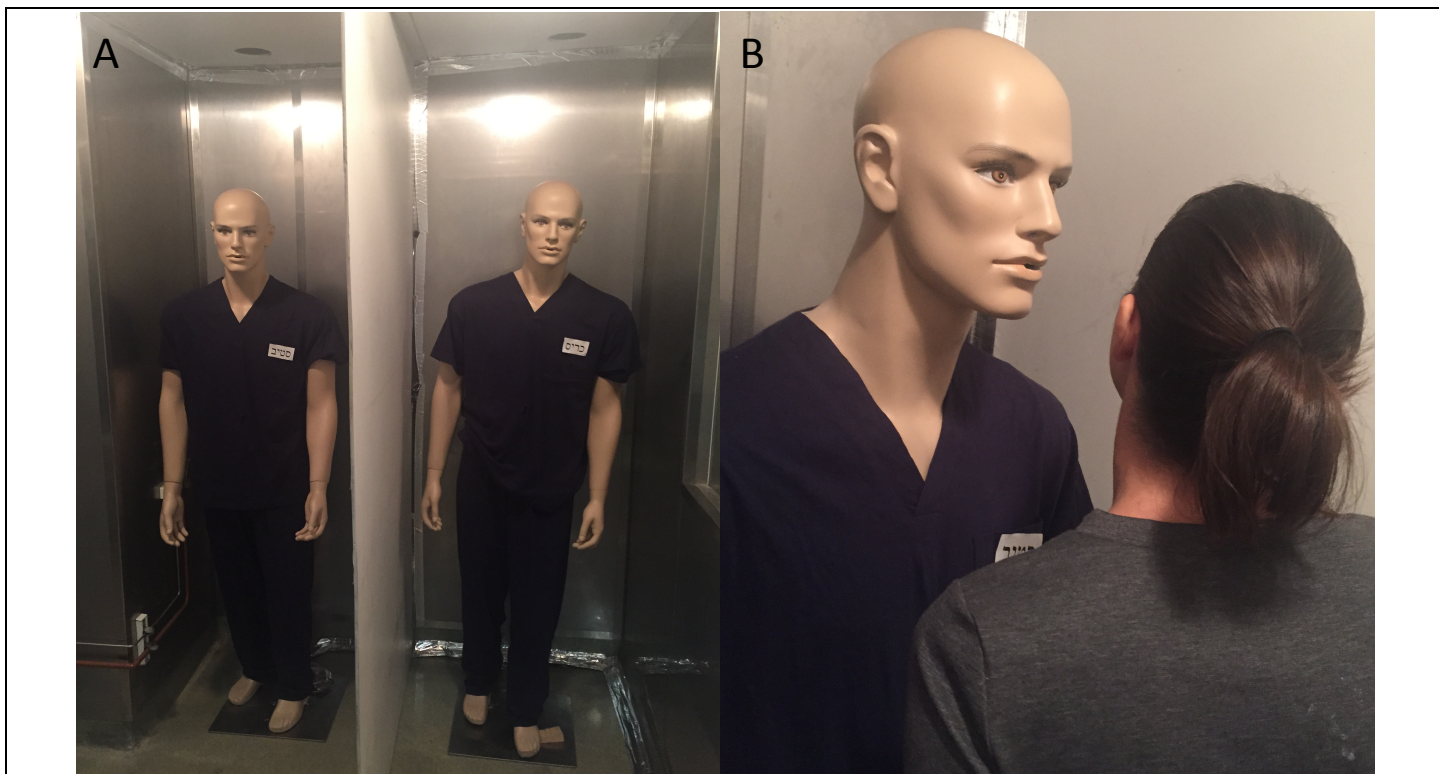
The Stroop task contained only one trial, and was thus not optimized for this analysis. Nevertheless, we observe a trend very similar to that seen in the Faces task, whereby the smell of fear increased arousal in TD but decreased it in ASD. An ANOVA applied to the EDA trace obtained over the ~40 sec task revealed no main effect (all $F_{1,33} < 0.3$, all $p > 0.6$), yet near significant interaction ($F_{1,33} = 4.0$, $p = 0.054$). This reflected that exposure to the smell of fear relative to control non-significantly increased EDA in TD (normalized Trough to Peak, Control = 0.94 ± 0.38 Nμs, Fear = 1.17 ± 0.40 Nμs, $t(19) = 1.6$, $p = 0.12$) yet exposure to the smell of fear relative to control non-significantly decreased EDA in ASD (normalized Trough to peak, Control = 1.15 ± 0.48 Nμs, Fear = 0.89 ± 0.47 Nμs, $t(14) = 1.3$, $p = 0.24$). All tests were two-tailed, all centers reflect mean, all error bars reflect SEM.



Supplementary Figure 5

Relation between ER-EDA and AQ

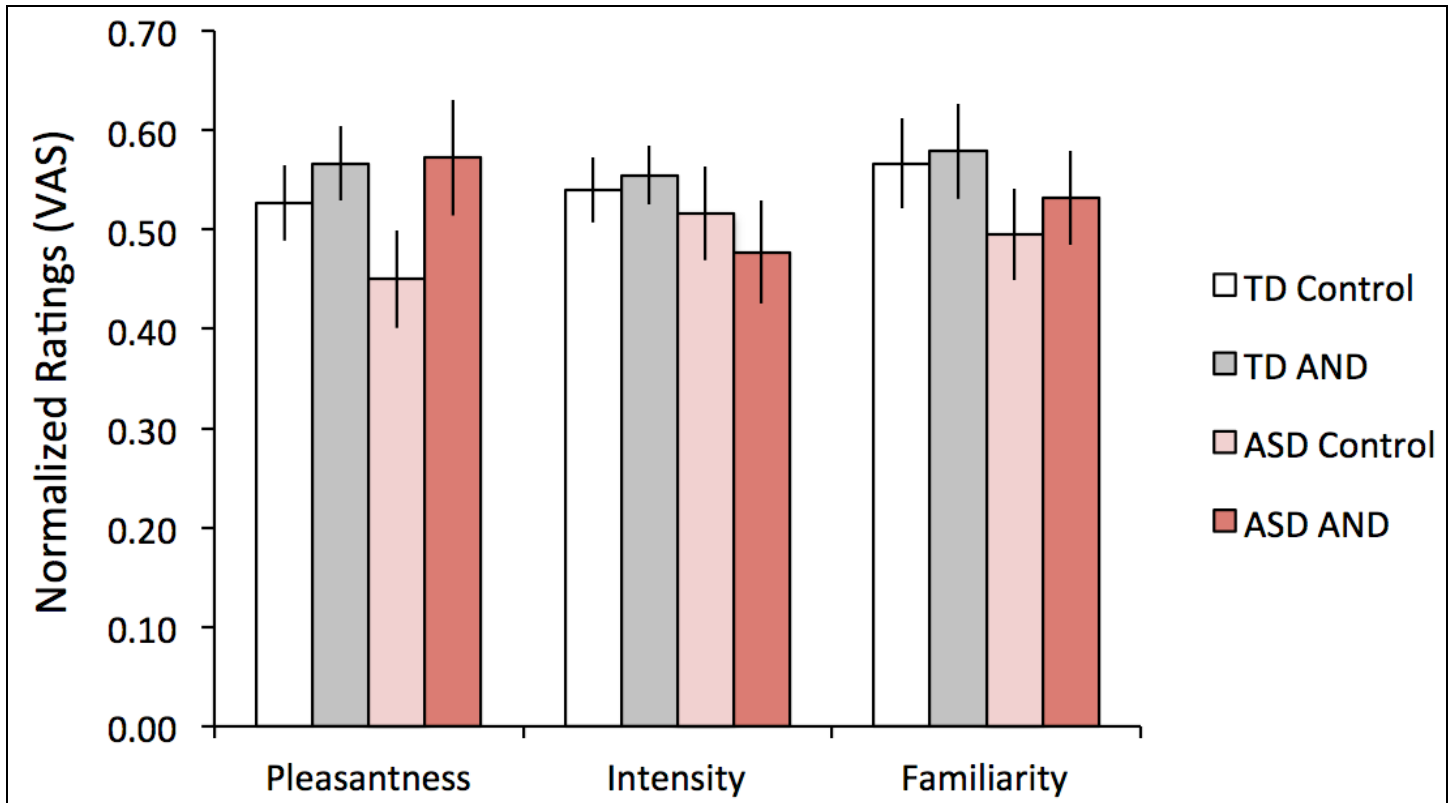
The relation between change in ER-EDA in the Faces task and AQ scores of all participants, TD (N =19) and ASD (n =17): $r = -0.33$, $p = 0.043$. This effect does not survive correction for the multiple comparisons involved.



Supplementary Figure 6

Manikin experimental set-up

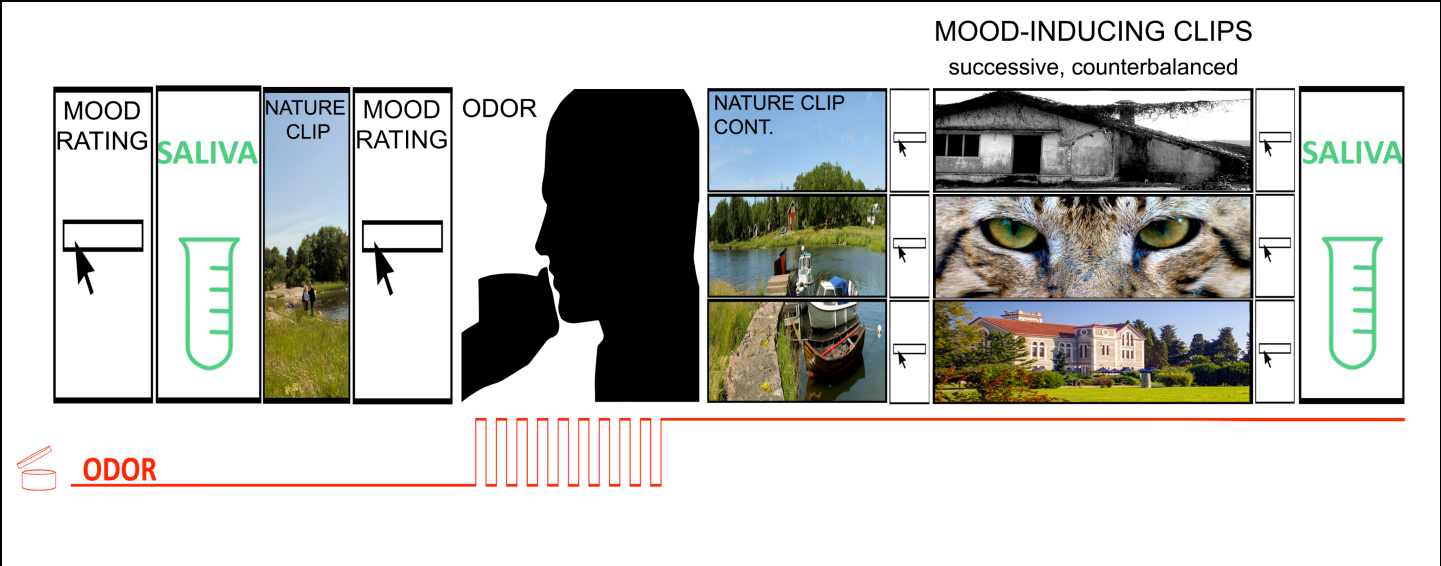
A. The actual location of the manikins in the experimental room. Note that a thin wall separates the two manikins, and each has a high-flow exhaust vent directly overhead. This design prevented contamination. Also note that each manikin is identified by nametag, which is switched across manikins across experiments. **B.** The typical posture of a participant interacting with the manikins. Because the manikin voice volume was very low, participants tended to put their ears up close to the manikin, thus putting their nose in line with the emitted chemosignal.



Supplementary Figure 7

Ratings of AND intensity pleasantness and familiarity

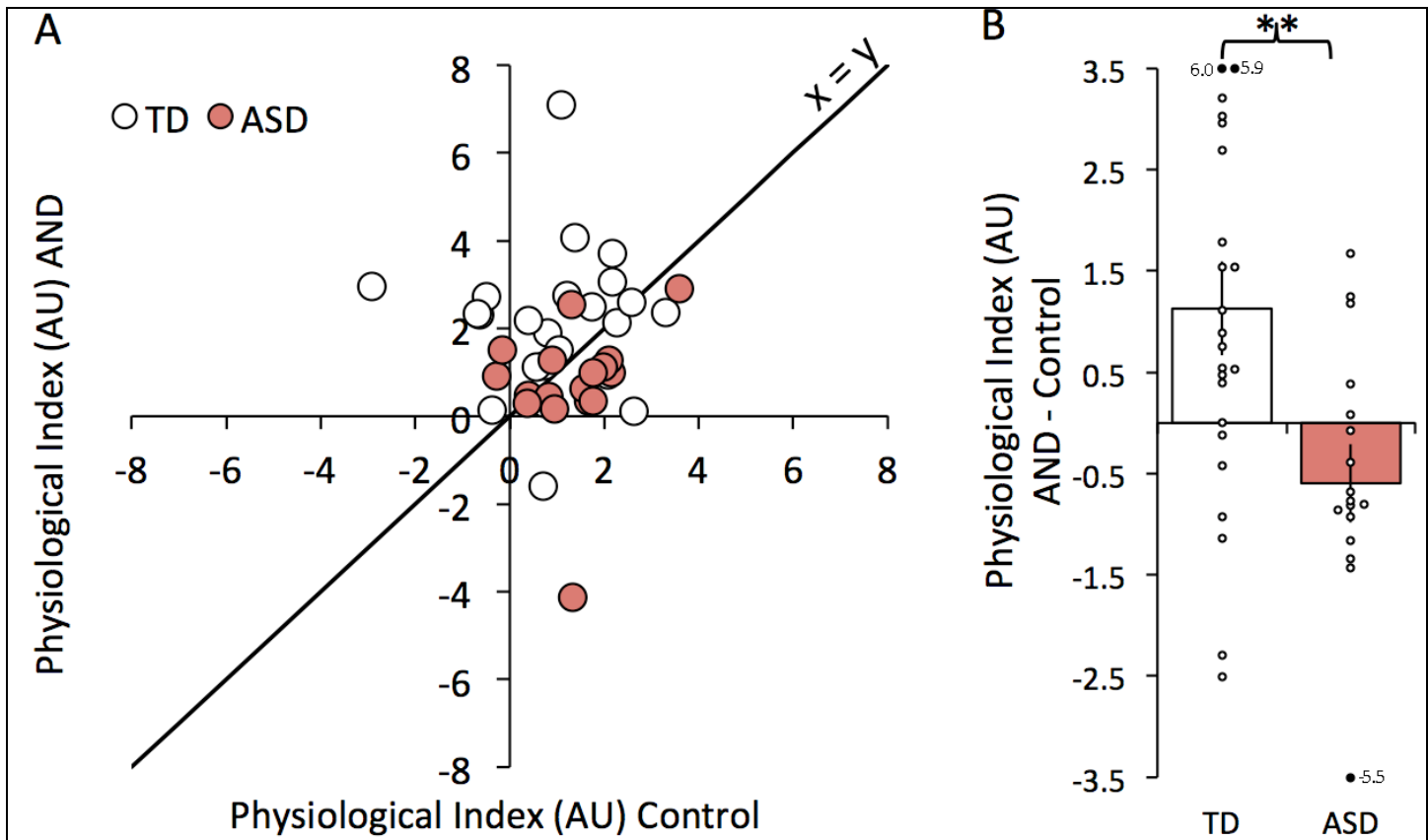
Mean $\pm$ SEM Ratings of stimulus intensity pleasantness and familiarity. Reported perception was similar for the clove carrier alone and for the clove carrier with AND in both TD ($n = 23$) and ASD ($n = 17$) groups. A two-way repeated-measures ANOVA on the ratings of stimulus intensity pleasantness and familiarity applied by the participants directly before the ensuing task, with factors of Condition (AND/control), Descriptor (Intensity/Familiarity/Pleasantness) and a categorical independent factor of Group (TD/ASD) revealed no main effect (all $F_{1,38} < 2.8$, all $p > 0.1$), and no interaction (all $F_{2,76} < 1.4$, all $p > 0.26$). In other words, verbally reported perception was similar for the clove carrier alone and for the clove carrier with AND in both TD and ASD. All tests were two-tailed, all centers reflect mean, all error bars reflect SEM.



Supplementary Figure 8

Experimental timeline with Androstadienone

Each participant tested on separate days at the same time, once with the chemosignal (AND) and once with control, counterbalanced for order, double-blind. All photos and graphics: Ofer Perl

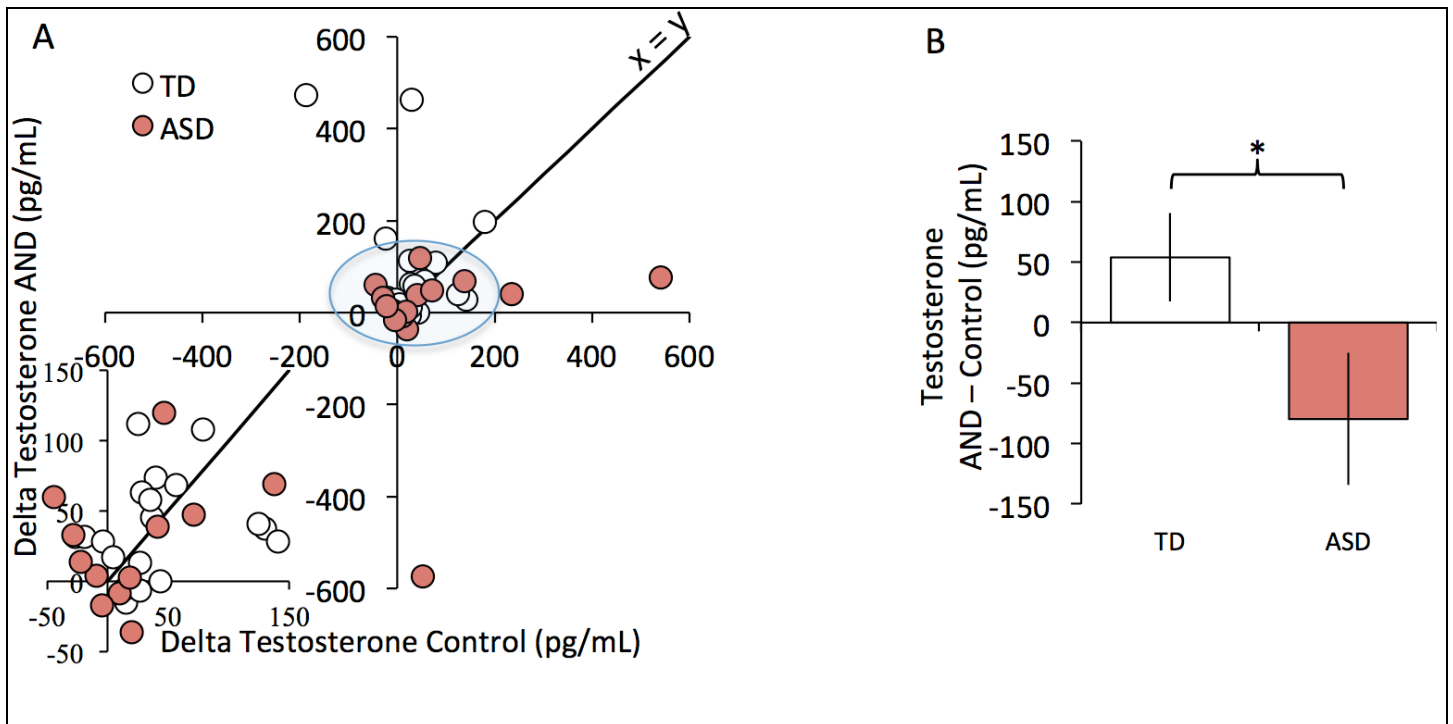


Supplementary Figure 9

AND increased arousal in TD but not in ASD

A. The influence of AND on ongoing changes in physiological index. Each circle (TD = clear, ASD = red) represents the change in physiological index of a single participant following AND (Y-axis) and Control (X-axis). The diagonal line reflects the unit slope line ($x=y$) ($n = 23$ TD, 17 ASD). **B.** Change in physiological index (summed results of the data in A). The composite physiological index combined the different physiological parameters (NS-EDA, Change in temperature and HRV) with equal weighting (1). An ANOVA with factors of Condition (AND/Control), Time (Before/After CP) and Group (ASD/TD) revealed a significant main effect of Time ($F_{1,38} = 57.8$, $p < 10^{-8}$, Before CP = -0.66 ± 1.3 AU, After CP = 0.66 ± 1.2 AU) and a significant interaction of Group X Condition X Time ($F_{1,38} = 7.8$, $p = 0.008$, non-parametric reanalysis: $U = 99$, $Z = 2.6$, $p = 0.008$, Cohen's $d' = 0.92$). This reflected a similar physiological index change across the experiment in TD and ASD in response to the control odor (TD = 0.97 ± 1.39 AU, ASD = 1.31 ± 0.95 AU, $t(38) = 0.85$, $p = 0.40$). Yet, the addition of undetected AND significantly reversed the drop in TD (Δ control = 0.97 ± 1.39 AU, Δ AND = 2.10 ± 1.69 AU, $t(22) = 2.5$, $p = 0.02$), but had no influence in ASD (Δ control = 1.31 ± 0.95 AU, Δ AND = 0.71 ± 1.46 AU, $t(16) = 1.6$, $p = 0.14$). All tests were two-tailed, all centers reflect mean, all error bars reflect SEM. ** = $p < 0.01$.

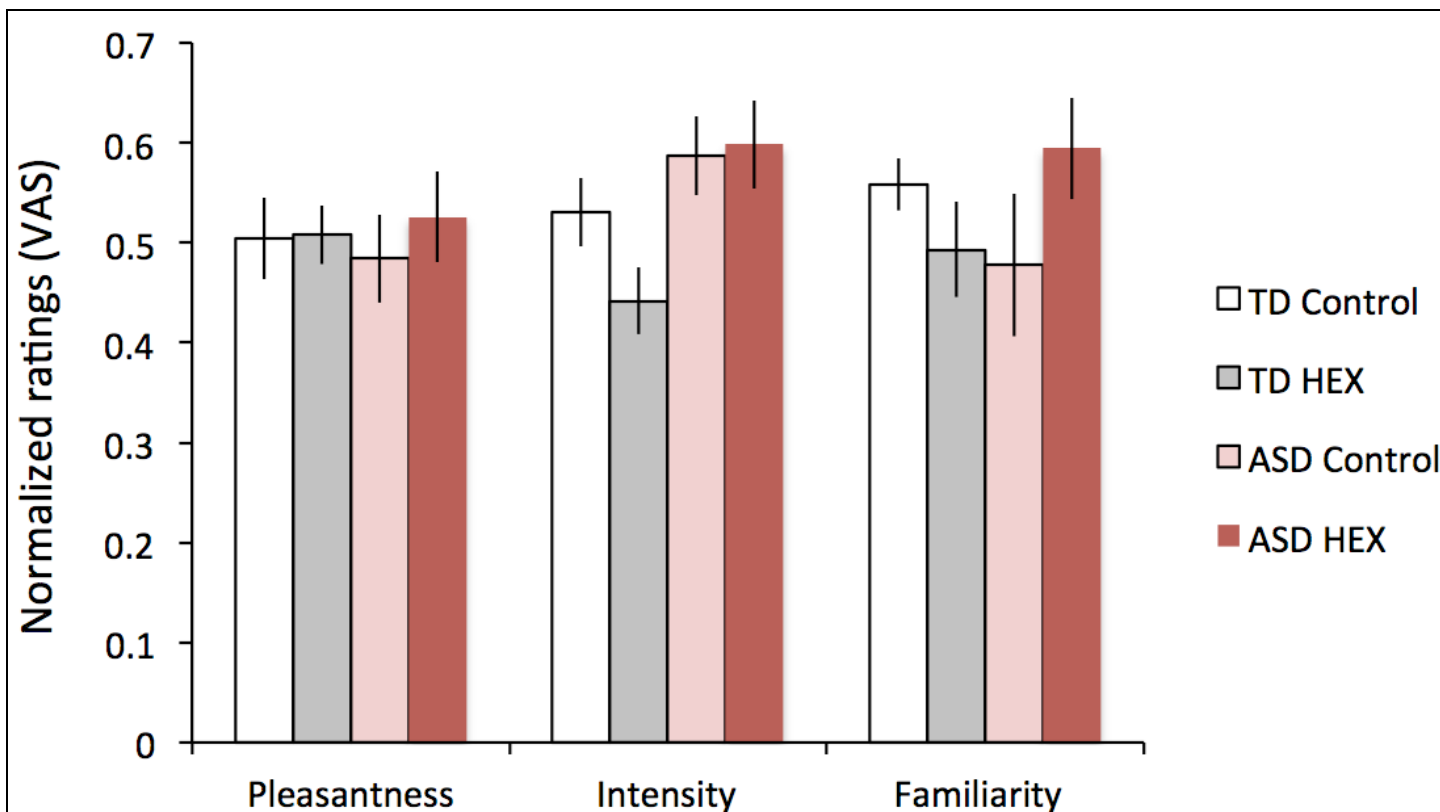
1. Bensafi, M., Brown, W. M., Khan, R., Levenson, B. & Sobel, N. Sniffing human sex-steroid derived compounds modulates mood, memory and autonomic nervous system function in specific behavioral contexts. *Behav Brain Res* 152, 11-22 (2004).



Supplementary Figure 10

Influence of AND on Testosterone

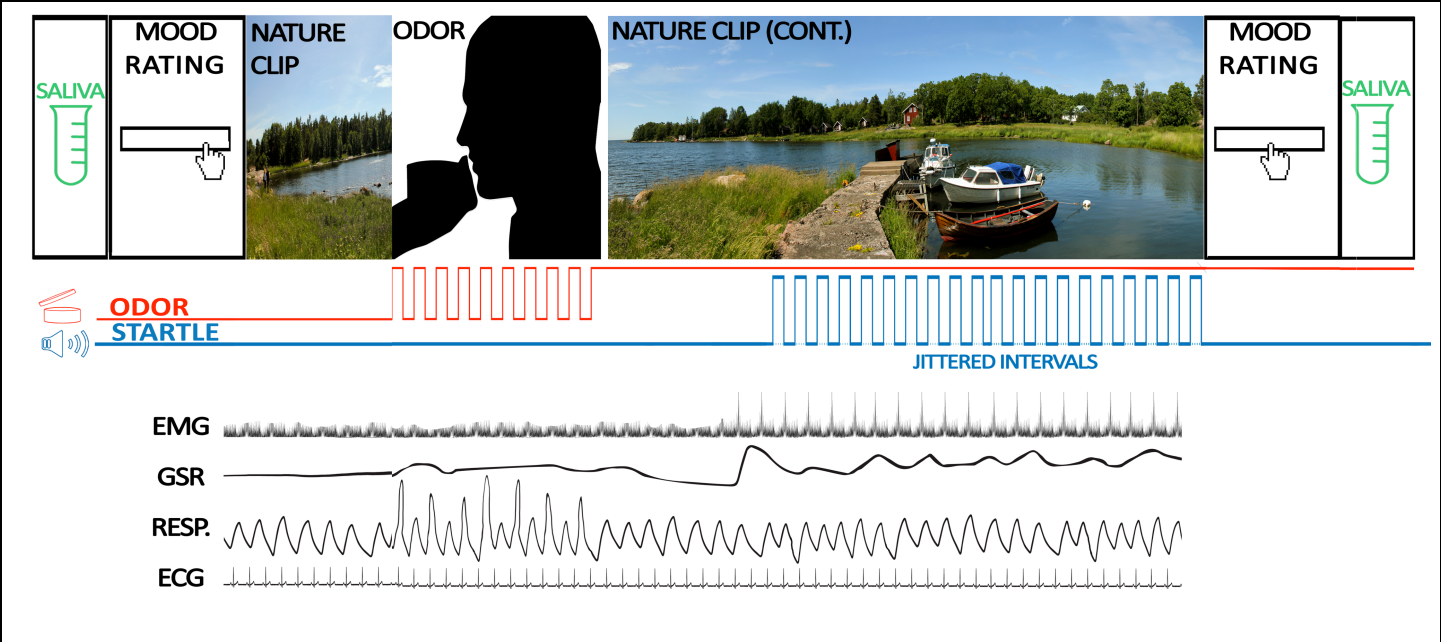
A. The influence of AND on testosterone of TD ($n = 23$, clear) and ASD ($n = 15$, red). (a) Each circle represents the testosterone ratio of a single participant following AND (Y-axis) and Control (X-axis). **B.** The delta testosterone. Measurements of testosterone in saliva revealed a pattern similar to that evidenced in physiological monitoring. An ANOVA with factors of Condition (AND/Control), Time (Start/End of the experiment) and Group (ASD/TD) revealed a main effect of Time ($F_{1,36} = 10.1$, $p = 0.003$, Testosterone concentration pg/mL Start = 168.7 ± 72.7 , End = 218.6 ± 87.5) and a significant interaction of Condition X Time X Group ($F_{1,36} = 4.8$, $p = 0.035$, Cohen's $d' = 0.72$). This reflected that in both TD and ASD testosterone similarly increased across the experiment. The addition of undetected AND slightly further increased testosterone in TD (Testosterone concentration, Δ Control = 35.0 ± 72.6 pg/mL, Δ AND = 87.9 ± 130.6 pg/mL, $t(22) = 1.5$, $p = 0.14$), but slightly decreased testosterone in ASD (Testosterone concentration, Δ Control = 73.5 ± 147.6 pg/mL, Δ AND = -8.7 ± 165.7 pg/mL, $t(14) = 1.5$, $p = 0.15$). This is consistent with the EDA results reported in the main text whereby AND increased arousal in TD and decreased it in ASD. All tests were two-tailed, all centers reflect mean, all error bars reflect SEM. * = $p < 0.05$.



Supplementary Figure 11

Ratings of HEX intensity pleasantness and familiarity

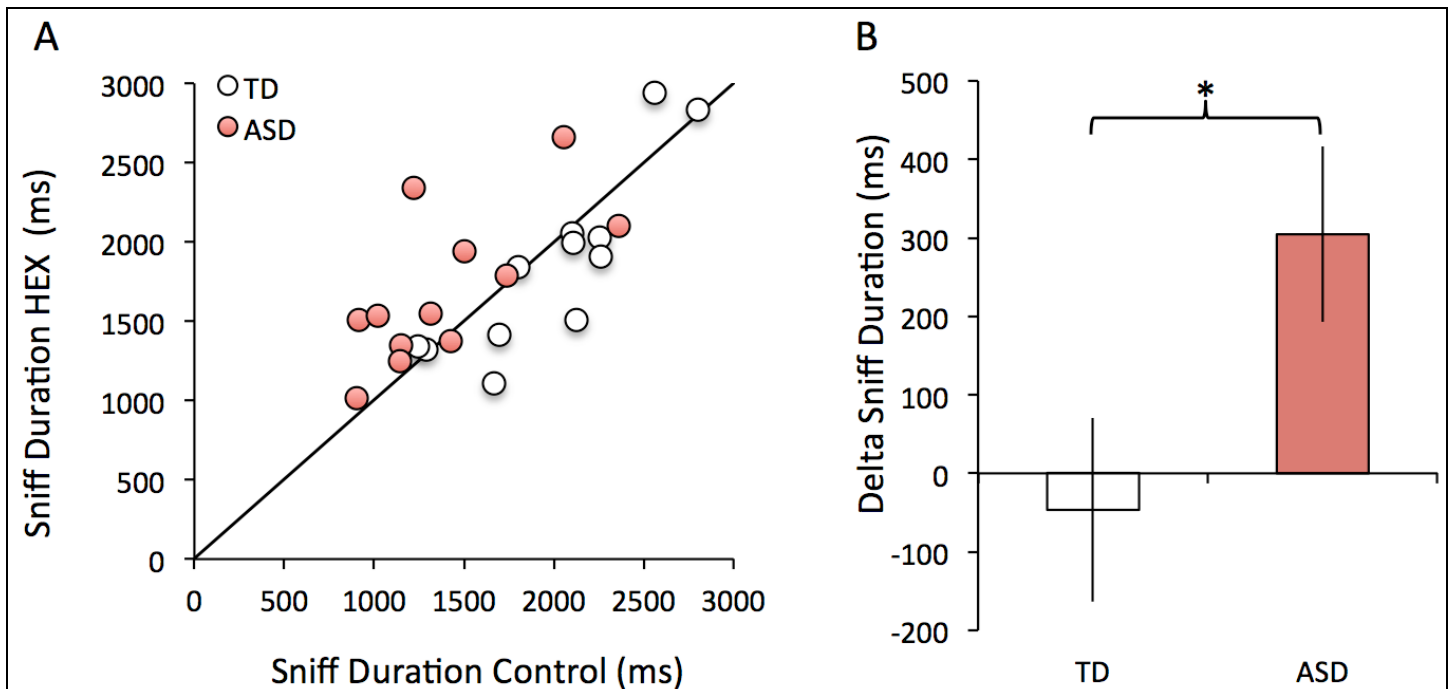
Mean $\pm$ SEM Ratings of stimulus intensity pleasantness and familiarity. Verbally reported perception was similar for the clove carrier alone and for the clove carrier with HEX in both TD ($n = 16$) and ASD ($n = 17$) groups. A two-way repeated-measures ANOVA on the ratings of stimulus intensity pleasantness and familiarity applied by the participants during the ensuing tasks with factors of Condition (HEX/control), Descriptor (intensity/familiarity/pleasantness) and a categorical independent factor of Group (TD/ASD). This revealed no main effects (all $F < 2.6$, all $p > 0.1$), no interaction of Descriptor X Group X Condition ($F_{2,62} = 0.9$, $p = 0.41$), and a trend towards an interaction of Group X Condition ($F_{1,31} = 3.5$, $p = 0.072$). In other words, verbally reported perception was similar for the clove carrier alone and for the clove carrier with HEX in both TD and high-function ASD. All tests were two-tailed, all centers reflect mean, all error bars reflect SEM.



Supplementary Figure 12

Experimental timeline with Hexadecanal

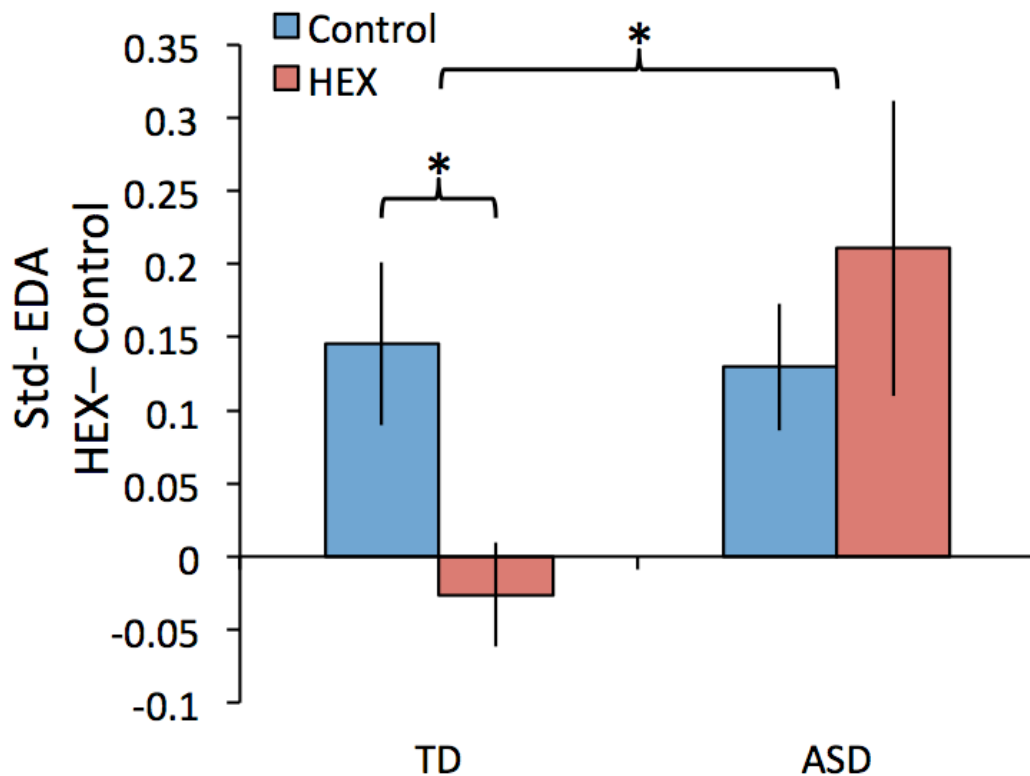
Each participant tested on separate days at the same time, once with the chemosignal (HEX) and once with control, counterbalanced for order, participants blind to condition. All photos and graphics: Ofer Perl



Supplementary Figure 13

Hex increased sniff duration in ASD but not in TD

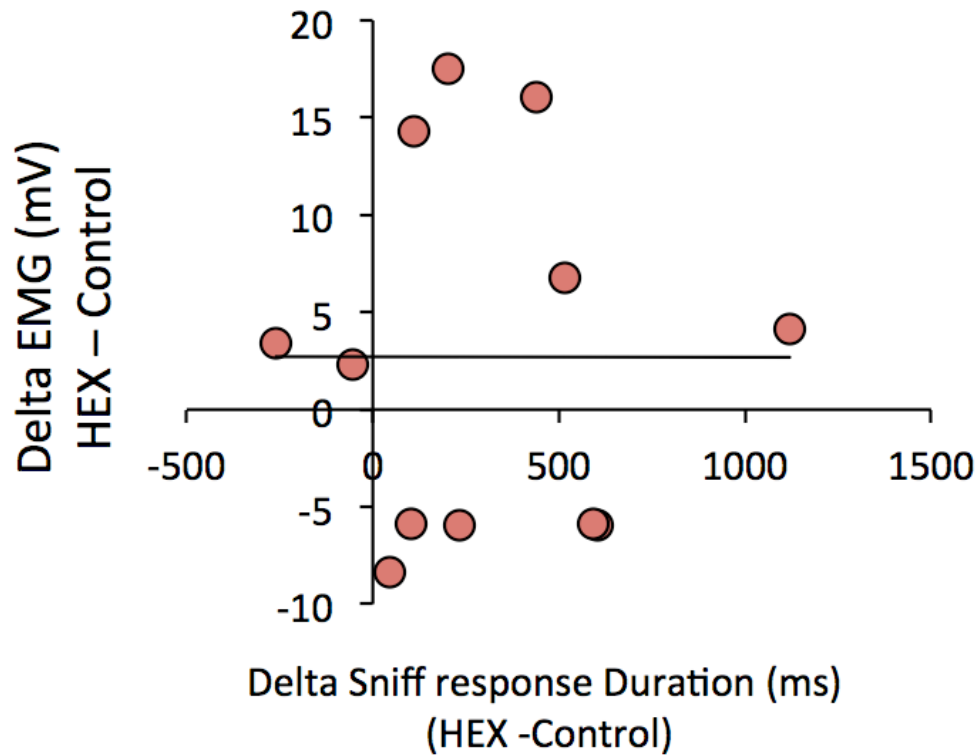
A. The influence of HEX on the sniff duration (ms) relative to control of TD ($n = 15$, white) and ASD ($n = 12$, red). Each circle represents mean sniff duration (ms) of a single participant following HEX (Y-axis) and Control (X-axis) **B.** Mean $\pm$ SEM Sniff duration. Because both ASD women (and two of the ASD men) refused to wear the nasal cannula, we compared the ASD group to the TD men only. A one way repeated-measures ANOVA applied to each of the standard sniff parameters (Volume/Duration/Peak), with conditions of Condition (HEX/Control) and categorical independent factor of Group (TD/ASD) revealed no main effects or interactions for Volume and Peak (all $F_{1,25} < 2.3$, $p > 0.14$), but for Duration we observed a significant main effect of Group ($F_{1,25} = 7.4$, $p = 0.011$), and a significant Condition X Group interaction ($F_{1,25} = 4.9$, $p = 0.036$, Cohen's $d' = 0.87$). This reflected that in ASD but not in TD mean sniff duration was longer for undetected HEX than for control (ASD: Control = 1396.5 ± 454.1 ms, HEX = 1701.7 ± 481.3 ms, $t(11) = 2.9$, $p = 0.014$, TD: Control = 2212.1 ± 617.1 ms, HEX = 2165.5 ± 864.8 ms, $t(14) = 0.41$, $p = 0.69$). This result further implies that the results of this study are not a reflection of an inability of the ASD brain to register the stimulus. The ASD brain differed between HEX and control as evidenced here in the longer duration sniffs, yet the autonomic impact in ASD was opposite that of TD. All tests were two-tailed, all centers reflect mean, all error bars reflect SEM. * = $p < 0.05$.



Supplementary Figure 14

NS-EDA decreased following HEX in TD but not in ASD

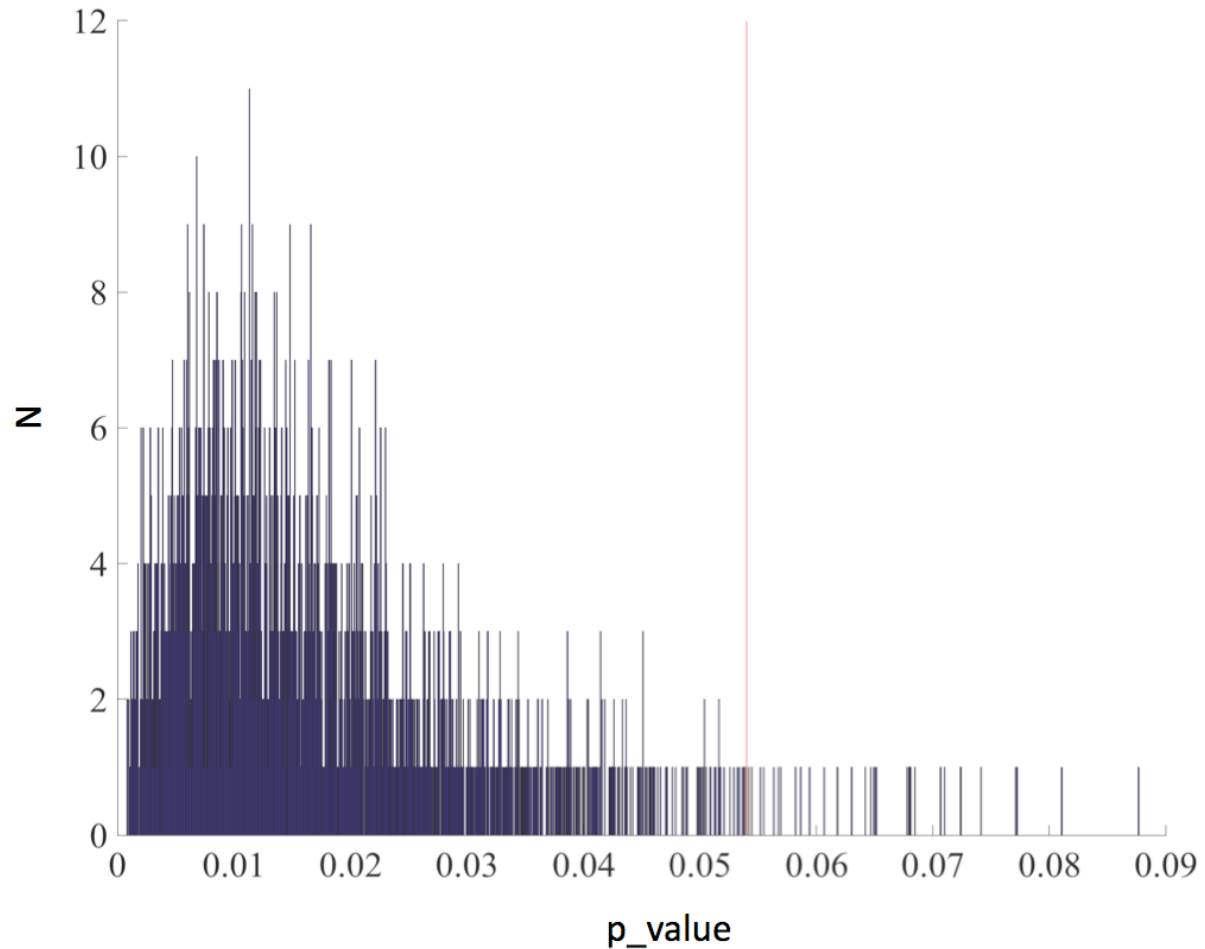
Mean ± SEM NS-EDA. Mean NS-EDA decreased following exposure to HEX (red) than for control (blue) in TD ($n = 17$) but not in ASD ($n = 15$). An ANOVA applied to the standard deviation in skin conductance (NS-EDA) revealed a significant main effect of Time ($F_{1,30} = 14.4$, $p < 10^{-4}$) and a significant interaction of Condition X Time X Group ($F_{1,30} = 4.7$, $p = 0.038$, Cohen's $d' = 0.75$) showing that HEX reduced NS-EDA in TD (Δ Control = $0.14 \pm 0.21 \mu s$, Δ HEX = $-0.03 \pm 0.13 \mu s$, $t(16) = 2.8$, $p = 0.012$) but had no influence on ASD (Δ Control = $0.13 \pm 0.16 \mu s$, Δ HEX = $0.21 \pm 0.38 \mu s$, $t(14) = 0.78$, $p = 0.45$). This ongoing measure is consistent with the startle response to imply an autonomic calming effect of HEX in TD but not ASD. All tests were two-tailed, all centers reflect mean, all error bars reflect SEM. * = $p < 0.05$.



Supplementary Figure 15

Startle effects not driven by sniffing

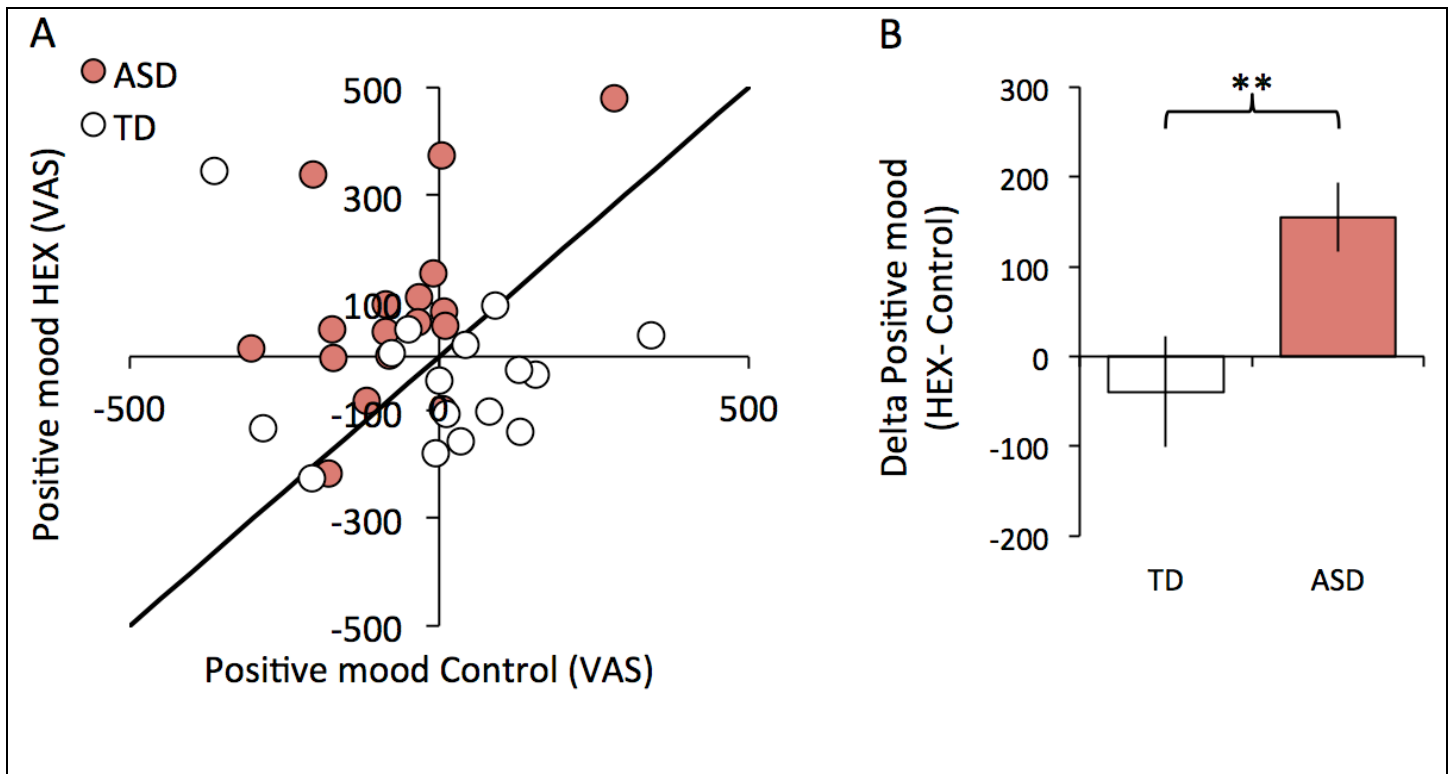
The extent of change in EMG as a function of the extent of change in sniff duration in ASD ($n = 12$). Given the altered sniffing for HEX one could raise the possibility that this is driving the startle response. To address this possibility, we tested for a relation between change in startle and change in nasal airflow in ASD, yet observed no relation ($r = 0.063$, $p = 0.85$).



Supplementary Figure 16

The distribution of p values following reiteratively removing 2 TD men and 2 ASD men from the startle analysis

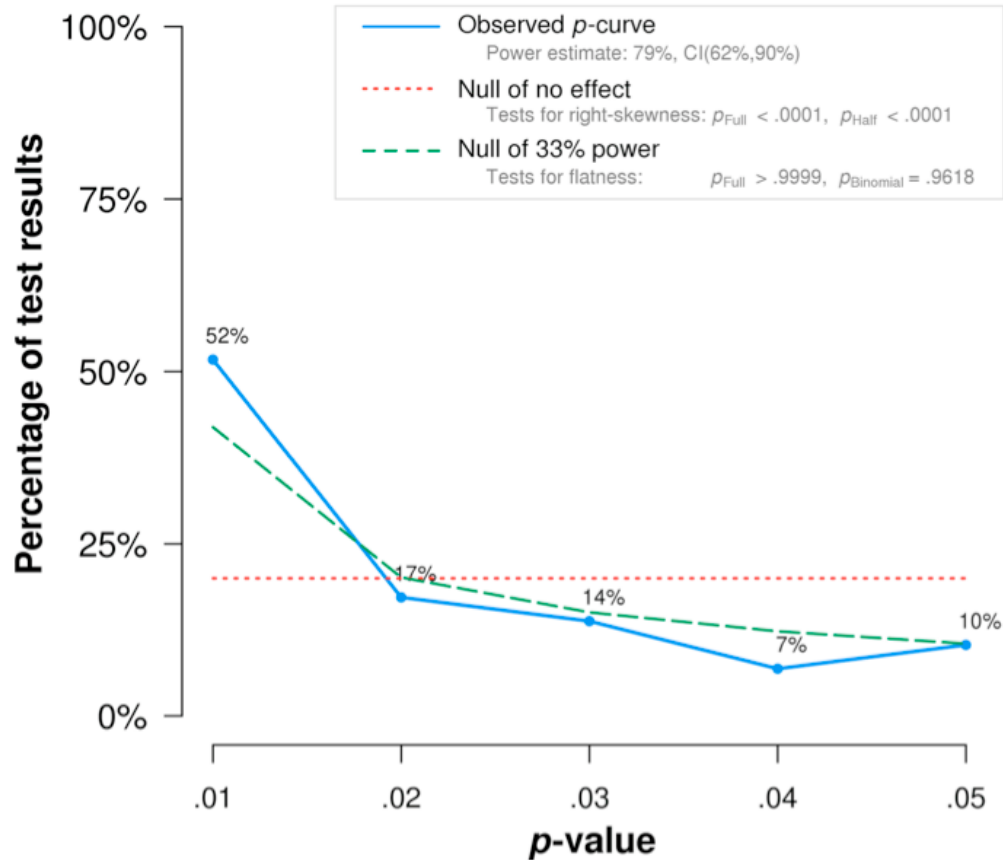
The red line represents the p value following removing the women (2 TD and 2 ASD) from the startle analysis. To ask whether the effects of HEX on startle were gender-specific we excluded the women from analysis of that one study (2 TD, 2 ASD). The ANOVA revealed a near significant interaction of Condition X Group ($F_{1,26} = 4.1$, $p = 0.054$). Notably, this analysis resulted in decreased significance compared with the original analysis that included all participants. In turn, this has now reduced our sample size by 4 participants. To estimate whether this was the effect of gender or the effect of sample size, we reiteratively removed 4 men (2 TD and 2 ASD) 1000 times, and tested the distribution of p values. We obtained a mean of $p = 0.016 \pm 0.012$, which is indeed significantly different from 0.054. Thus, we conclude that the effect was in the same direction in men and women, but stronger in women that participated in the experiment.



Supplementary Figure 17

HEX impacted self-reported mood in ASD but not TD

A. The influence of HEX on the positive mood rating relative to control of TD ($n = 16$, white) and ASD ($n = 17$, red). Each circle represents the positive mood rating of a participant following HEX (y-axis) and Control (x-axis). **B.** The positive mood of the ASD participants increased significantly compared with TD ($F_{1,31} = 7.8$, $p = 0.009$). A two-way repeated-measures ANOVA with conditions of Condition (Hexadecanal/Control), Time (Beginning/End of the experiment) and a categorical independent factor of Group (TD/ASD) was applied to each of the mood subscales (Positive/Negative low/Negative high/Sexual arousal). The ANOVA on Positive mood revealed a significant interaction of Condition X Time X Group ($F_{1,31} = 7.8$, $p = 0.009$, Cohen's $d' = 0.97$) (Supplementary Figure 19). This reflected a significant increase in the positive mood of the ASD group alone following exposure to undetected HEX (ASD: $\Delta \text{Hex} = 86.3 \pm 174.1$, $\Delta \text{Control} = -68.5 \pm 128.1$, $t(16) = 4.1$, $p = 0.0008$, TD: $\Delta \text{Hex} = -37.1 \pm 136.9$, $\Delta \text{Control} = 2.0 \pm 174.4$, $t(15) = 0.65$, $p = 0.52$). ANOVAs on the other subscales revealed no interaction or main effect that involved Condition (all $F < 2.5$, all $p > 0.1$). In other words, the influence of HEX on self-reported mood in this study was limited to one subscale alone. Although this again points to dissociable impact of HEX in TD and ASD, we are cautious in interpretation of this result. This is because unlike the psychophysiological results that emerged again and again throughout all experiments in this study, this self-reported mood result is the only significant case throughout the study. Although its extent is such that it in fact survives correction for multiple comparisons following from the above, we nevertheless retain caution in interpretation. All tests were two-tailed, all centers reflect mean, all error bars reflect SEM. ** = $p < 0.01$.

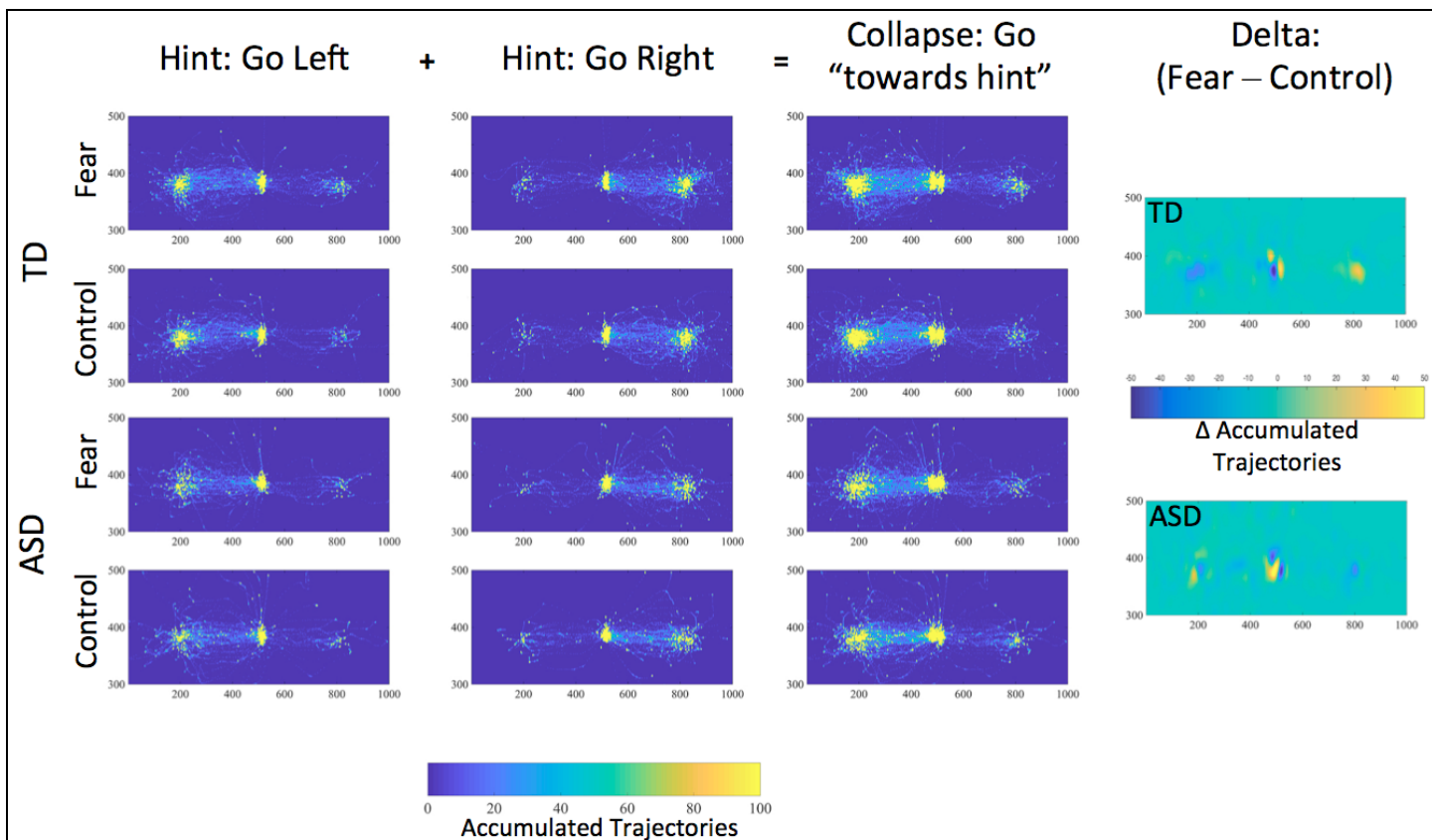


Supplementary Figure 18

The distribution of the reported p-values throughout the manuscript

We take a meta-analysis approach to the experiments we conducted. In order to verify that our reported p values in the manuscript are not selective and to test whether our findings across experiments reflect true effects, we applied p-curve analysis (1). We found that the p-curve of all our experiments combined is significantly right skewed (Full p-curve, $Z = -7.24$, $p < 10^{-5}$, Half p-curve, $Z = -5.88$, $p < 10^{-5}$), indicating overall evidential value of the reported results.

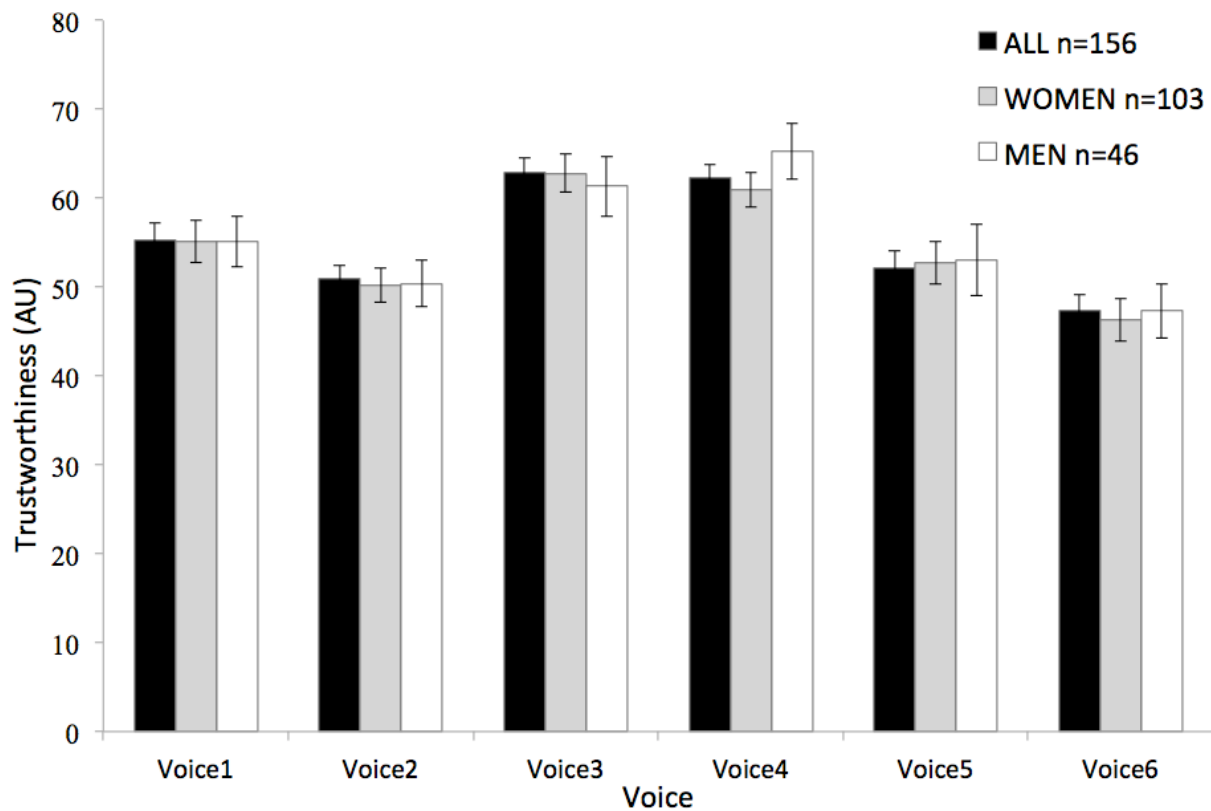
1. Simonsohn, U., Nelson, L. D. & Simmons, J. P. P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science* 9, 666-681 (2014).



Supplementary Figure 19

Mouse trajectory presentation

The following heat maps depict an accumulation of all mouse locations on the task screen within 1000 ms from trial onset. Each of the 12 sub-plots corresponds to mouse movement of a specific experimental group. Color code denotes the number of accumulated counts of a certain pixel coordinate being "visited" by the mouse cursor throughout all trials. One can observe fine single-trial trajectories traveling from the center to arrive at the expected location of the target. From top to bottom: trials of TD participants approaching the Smell of Fear manikin, trials of TD participants approaching the control manikin, trials of ASD participants approaching the Smell of Fear manikin, trials of ASD participants approaching the control manikin. Horizontal division into sub plots is based on the hint given by the manikin. From left to right: Hint to left target, hint to right target and a superposition of both hints directions created by adding of left target and a mirror image of right target to result in a "go towards target" disregarding screen orientation. Lastly, in the rightmost panel is the comparison of response to Smell of fear vs. control from both the TD group (top) and ASD group (bottom). Color code denotes the delta of accumulated trajectories per group such that "hot" colors denote a higher number of trajectories that went through the pixel coordinate while "cold" colors denote less "visits" of a certain pixel coordinate.



Supplementary Figure 20

Manikin voice selection

The results of an online study comparing the trustworthiness associated with 6 different potential voices for the manikins. All centers reflect mean, all error bars reflect SEM (n = 198).

Code	GENDER	Participated in Experiment						AQ	Age Range (years)
		FEAR FACES	FEAR TRUST	AND	HEX	FEAR RATINGS	DISCRIMINATION		
TD1	M				+			17	25
TD2	M				+			20	27
TD3	M				+			16	25
TD4	M				+			19	24
TD5	M			+		+	+	18	29
TD6	M			+				10	29
TD7	M					+		10	26
TD8	M				+		+	15	26-30
TD9	M		+					8	34
TD10	M			+				13	25
TD11	M			+			+	28	24
TD12	M			+				19	24
TD13	M			+				20	28
TD14	M			+				13	24
TD15	M			+				10	25
TD16	M			+				9	42
TD17	M			+				27	26
TD18	M	+	+	+		+	+	16	24-26
TD19	M	+		+		+	+	16	24-27
TD20	M		+					19	27
TD21	M				+	+		21	28
TD22	M	+						16	31
TD23	M					+	+	24	28
TD24	M				+			10	26
TD25	M				+			18	28
TD26	M	+		+				18	25-26
TD27	M			+			+	19	24
TD28	M			+			+	15	26
TD29	M			+			+	15	25
TD30	M						+	23	35
TD31	M						+	14	24
TD32	M						+	20	32
TD33	M						+	13	33
TD34	M						+	13	27
TD35	M	+					+	13	25-27
TD36	M						+	11	24
TD37	M						+	18	24
TD38	M	+						21	25
TD39	M	+	+	+				22	27-28
TD40	M		+					17	29
TD41	M	+		+		+		16	27-28
TD42	M		+					15	27
TD43	M					+		16	33
TD44	M					+		14	26
TD45	M	+	+					18	28
TD46	M	+	+					8	29
TD47	M	+		+		+		21	30-31
TD48	M			+				15	28
TD49	M					+		20	25
TD50	M					+		17	28
TD51	M					+		11	24
TD52	M			+				15	24
TD53	M					+		26	19
TD54	M			+		+		10	29
TD55	M			+				8	32
TD56	M	+	+					16	32-33
TD57	M	+	+					26	26
TD58	M		+					21	25
TD59	M		+					20	26
TD60	M	+						21	25
TD61	M	+						15	25
TD62	M	+						10	31
TD63	M	+						15	27
TD64	M	+						19	27
TD65	M	+	+					23	35
TD66	M	+	+					11	23
TD67	M		+					10	26
TD68	M		+					19	28
TD69	M		+					27	26
TD70	M		+					17	20
TD71	M		+					19	28
TD72	M		+					17	29
TD73	M				+			15	28
TD74	M				+		+	14	33
TD75	M				+			14	29
TD76	M				+			13	29
TD77	M				+			21	29
TD78	M				+			19	27
TD79	f				+			12	24
TD80	M				+			19	25
TD81	f				+			11	24

Supplementary Table 1. TD participants

This table details the gender, age, AQ of the TD participants, as well as their participation in the different experiments and rating tasks. Note that age appears as a range because these different experiments were conducted across about four years.