

Supplemental Methods

Stimuli for IATs

Gender IAT: Male names ('Ben', 'John', 'Daniel', 'Paul', 'Jeffrey', 'Michael') and female names ('Julia', 'Michelle', 'Anna', 'Emily', 'Rebecca', 'Lisa') were paired with family words ('Home', 'Parents', 'Children', 'Family', 'Marriage', 'Wedding') and career words ('Management', 'Corporation', 'Salary', 'Office', 'Business', 'Career').

Race Evaluation IAT: Greyscale photographs of 6 African American (3 male, 3 female) and 6 Caucasian (3 male, 3 female) faces with neutral facial expressions were paired with positive words ('Joy', 'Love', 'Wonderful', 'Pleasure', 'Glorious', 'Happy') and negative words ('Agony', 'Terrible', 'Horrible', 'Nasty', 'Evil', and 'Awful').

Race Stereotype IAT: The same images from the Race Evaluation IAT were paired with words depicting intelligence (intellectual words: 'Math', 'Brainy', 'Educated', 'Scientist', 'College', 'Genius') and athleticism (Athletic words: 'Athletic', 'Boxing', 'Basketball', 'Run', 'Jump', 'Football').

Shoe IAT: Grayscale photographs of 6 Dress shoes and 6 Sneakers were paired with words depicting business ('Merger', 'Portfolio', 'Desk', 'C.E.O.', 'Investor', 'Cubicle) and sports ('Game', 'Ball', 'Score', 'Championship', 'Running', 'Exercise').

Flower IAT: Positive words and negative words (as above) were paired with Flowers ('Daisy', 'Rose', 'Tulip', 'Daffodil', 'Lilac', 'Lily') and insects ('Roach', 'Ant', 'Spider', 'Fly', 'Termite', 'Wasp').

Supplemental Results

Null Hypothesis Statistical Testing (NHST) (IAT tests)

Group x Social Content: To enable comparison with previous work using NHST, we submitted the data to a mixed 2 (Group: ASD, CT) by 2 (Social vs. Nonsocial IAT) repeated-measures ANOVA. The ANOVA revealed a significant main effect of Group, $F(1, 60) = 16.30, p < .001, \eta^2 = 0.21$, indicating that the ASD group showed an overall lower IAT effect than controls. In addition, a main effect of Social Content, $F(1, 60)=64.88, p<.001, \eta^2 = 0.52$, reflected that IAT effects were overall larger for the Nonsocial tests (Flowers, Shoes) than for the Social tests (Race Eval, Race Stereo, Gender Stereo). However, there was no Diagnosis x Social Content interaction, $F(1, 60) = 2.84, p=.10, \eta^2 = 0.05$, suggesting that the effect of group did not differ as a function of social content.

Standard Deviation: Regression Analysis

To investigate whether differences in RT variability could account for our pattern of results, we regressed *mean overall SD* (x) (for each Condition and Subject) against *IAT D* (y), for all participants (ASD and CTLs). The residuals from this regression were then used to compute group differences in IAT effects (i.e. with standard deviation regressed out). This analysis revealed the same pattern of results as before: the effect size estimate for the group difference was again large for nonsocial IATs ($d=1.03, 95\% \text{ CI } [.57, 1.57]$), and medium for social IATs ($d=.59, 95\% \text{ CI } [.07, 1.17]$), with 95% CIs in each case excluding zero. Again, the between-groups effect sizes for social and nonsocial IATs did not differ from each other (overlapping 95% CIs).

RT: Regression Analysis

Similar regressions were run to rule out the possibility that slower responding in ASD could account for their slightly weaker IAT effects relative to CTL.

(1) *Mean overall RT regressed out:* This analysis revealed the same pattern of results as the main analysis: the effect size estimate for the group difference was again large for nonsocial IATs ($d=1.05$, 95% CI [.58, 1.60]), and medium for social IATs ($d=.59$, 95% CI [.09, 1.17]), with 95% CIs in each case excluding zero.

(2) *Mean congruent RT regressed out:* This analysis revealed the same overall pattern of results as before but with weaker between-groups effect sizes: the effect size estimate for the group difference was medium for nonsocial IATs ($d=.70$, 95% CI [.20, 1.24]), and small for social IATs ($d=.23$, 95% CI [-.29, .76]), with 95% CIs in the latter case failing to exclude zero.

(3) *Mean incongruent RT regressed out:* This analysis revealed the same pattern of results as the main analysis with even stronger between-group differences: the effect size estimate for the group difference was large for nonsocial IATs ($d=1.12$, 95% CI [.66, 1.65]), and medium/large for social IATs ($d=.76$, 95% CI [.25, 1.41]), with 95% CIs in each case excluding zero.

Supplemental Table 1. First-run analysis of IAT effects (analyzing the first nonsocial and nonsocial IATs administered). IAT D effects (means and bootstrapped 95% CIs), and between groups effect sizes (Cohen's d) in bold.

	All trials		First-run	
	ASD	CTL	ASD	CTL
nonSocial	.53 [.40, .63]	.79 [.72, .85]	.50 [.40, .62]	.83 [.74, .91]
Social	.29 [.19, .39]	.42 [.35, .48]	.40 [.21, .57]	.53 [.44, .63]
ASDvsCTL_Nonsocial_ES	1.03 [.57, 1.58]		1.18 [.66, 1.86]	
ASDvsCTL_Social_ES:	.56 [.05, 1.14]		.35 [-.17, .89]	

Supplemental Table 2. IAT sequence of blocks (example for a Flower-Insects IAT with congruent condition first, incongruent condition second)

Block	No. of Trials	Function	Items assigned to left-key response	Items assigned to right-key response
1	20	Practice	Flowers	Insects
2	20	Practice	Pleasant words	Unpleasant words
3	20	Practice	Flowers + Pleasant words	Insects + Unpleasant words
4	40	Test	Flowers + Pleasant words	Insects + Unpleasant words
5	40	Practice	Insects	Flowers
6	20	Practice	Insects + Pleasant words	Flowers + Unpleasant words
7	40	Test	Insects + Pleasant words	Flowers + Unpleasant words

Supplemental Table 3. IAT effects and their correlations with AQ/BAPQ scores for the MTurk (online) sample. Data presented for Whole Sample, White participants only, male participants only, and female participants only. Values represent mean D effects and Pearson r correlations (italicized) with AQ and BAPQ scores. Brackets contain bootstrapped 95% confidence intervals.

	Whole Sample (N=342)	White only (n=287)	Male only (n=183)	Female only (n=159)
Flowers/Insects	D = .76 [.71, .81]	D = .76 [.70, .81]	D = .66 [.58, .73]	D = .85 [.78, .91]
<i>AQ</i>	<i>r(340) = -.12</i> [-.28, .04]	<i>r(285) = -.14</i> [-.30, .04]	<i>r(181) = -.14</i> [-.37, .11]	<i>r(157) = -.07</i> [-.26, .15]
<i>BAPQ</i>	<i>r(340) = -.19</i> [-.34, -.03]	<i>r(285) = -.18</i> [-.34, -.002]	<i>r(181) = -.17</i> [-.41, .10]	<i>r(157) = -.16</i> [-.36, .06]
Shoes	D = .76 [.71, .80]	D = .76 [.71, .80]	D = .78 [.72, .83]	D = .73 [.66, .81]
<i>AQ</i>	<i>r(340) = .01</i> [-.14, .14]	<i>r(285) = .03</i> [-.13, .17]	<i>r(181) = .003</i> [-.20, .18]	<i>r(157) = -.01</i> [-.21, .18]
<i>BAPQ</i>	<i>r(340) = .05</i> [-.10, .18]	<i>r(285) = .02</i> [-.14, .17]	<i>r(181) = .04</i> [-.16, .22]	<i>r(157) = .03</i> [-.16, .22]
Race Eval	D = .44 [.38, .50]	D = .47 [.40, .53]	D = .49 [.39, .57]	D = .41 [.33, .48]
<i>AQ</i>	<i>r(340) = .04</i> [-.15, .21]	<i>r(285) = -.001</i> [-.21, .20]	<i>r(181) = .10</i> [-.23, .36]	<i>r(157) = -.03</i> [-.25, .19]
<i>BAPQ</i>	<i>r(340) = -.02</i> [-.18, .15]	<i>r(285) = -.01</i> [-.21, .18]	<i>r(181) = .08</i> [-.19, .36]	<i>r(157) = -.12</i> [-.32, .09]
Race Ster	D = .32 [.27, .38]	D = .32 [.27, .38]	D = .36 [.29, .42]	D = .27 [.19, .35]
<i>AQ</i>	<i>r(340) = .03</i> [-.11, .18]	<i>r(285) = .03</i> [-.14, .18]	<i>r(181) = -.07</i> [-.25, .11]	<i>r(157) = .17</i> [-.09, .40]
<i>BAPQ</i>	<i>r(340) = .13</i> [-.01, .26]	<i>r(285) = .11</i> [-.04, .25]	<i>r(181) = .03</i> [-.15, .20]	<i>r(157) = .25</i> [-.001, .43]
Gender	D = .37 [.32, .41]	D = .37 [.32, .41]	D = .35 [.29, .41]	D = .39 [.31, .46]
<i>AQ</i>	<i>r(340) = .10</i> [-.05, .23]	<i>r(285) = .09</i> [-.07, .24]	<i>r(181) = .17</i> [-.02, .34]	<i>r(157) = -.01</i> [-.28, .21]
<i>BAPQ</i>	<i>r(340) = .08</i> [-.05, .22]	<i>r(285) = .06</i> [-.08, .21]	<i>r(181) = .16</i> [-.03, .34]	<i>r(157) = -.02</i> [-.21, .15]

Supplemental Table 4. Demographic information (values represent means \pm standard deviations) for participant groups from a previous pilot study. Pilot ASD: participants with a diagnosis of ASD tested in the laboratory. Pilot CTL: controls tested in the laboratory.

	Pilot ASD (n=8)	Pilot CTL (n=27)
Age (yrs)	29.88 \pm 10.89	32.96 \pm 9.20
FSIQ	106.75 \pm 6.20	113.33 \pm 8.70
PIQ	97.88 \pm 8.59	110.81 \pm 10.51
VIQ	117.38 \pm 12.29	112.67 \pm 8.42
ADOS A (Cut-off 3/2)	4.13 \pm 1.73	--
ADOS B (Cut-off 6/4)	7.63 \pm 3.20	--
ADOS A+B (Cut-off 10/7)	11.75 \pm 4.74	--

Supplemental Table 5. IAT D effects for the combined samples (pilot participants plus participants from main study), for the three IAT tests that were common to both studies. One CTL participant from the pilot study did not complete the Race Evaluative IAT. Values are means and [bootstrapped 95% CIs].

	ASD combined (n=35)	CTL combined (n=65)
Flow/Ins IAT	0.53 [0.43, 0.63]	0.76 [0.67, 0.83]
Race Eval IAT	0.35 [0.20, 0.49]	0.47 [0.38, 0.56]
Gender IAT	0.27 [0.16, 0.40]	0.41 [0.34, 0.48]
nonSocial	0.53 [0.43, 0.63]	0.76 [0.67, 0.83]
Social	0.31 [0.21, 0.42]	0.44 [0.38, 0.50]

Supplemental Table 6: Explicit scale data (means \pm standard deviations) for the online (MTurk sample). MRS = Modern Racism Scale; IMS = Internal motivation to respond without prejudice; EMS = External motivation to respond without prejudice; Race Semantic = semantic differential scores for Black/White+Good/Bad; Flo/Ins Semantic = semantic differential scores for Flowers/Insects+Good/Bad; Shoe Semantic = semantic differential scores for Dress Shoes/Sneakers+Sports/Business; MSS = Modern Sexism Scale; Gender/Occupation = reported associations between Male/Female+Career/Family Occupation; Race/Occupation = reported associations between Black/White+Mental/Physical Occupation (Race Occ.)

Scale	Stereotype Group		Evaluative Group	
	Count	Mean \pm SD	Count	Mean \pm SD
MRS	178	12.65 \pm 5.05	164	13.04 \pm 5.47
IMS	178	7.03 \pm 2.09	164	7.19 \pm 1.94
EMS	178	4.02 \pm 2.21	164	4.01 \pm 2.13
Race Semantic	178	1.28 \pm 5.23	164	.63 \pm 4.93
Flo/Ins Semantic	--	--	164	14.12 \pm 6.88
Shoe Semantic	178	3.79 \pm .69	--	--
MSS	178	21.43 \pm 5.01	--	--
Gender/Occupation	178	3.41 \pm .40	--	--
Race/Occupation	165	3.55 \pm .50	--	--

Supplemental Table 7: Correlations (Pearson r 's) between ADOS scores and raw RT differences (incongruent minus congruent blocks) on the IAT, for the ASD group (n=27-35 depending on task). Each cell contains Pearson r and bootstrapped 95% CI. Raw Scores (A+B) = sum of raw scores on the Communication (A) and Social (B) subscales of the ADOS; Old Algorithm A+B = scores from Communication (A) + Social (B) scoring algorithm; New Algorithm (SA) = scores from Social Affect algorithm on revised ADOS; and New Algorithm Severity Score (SA) = calibrated Social Affect severity score (scores standardized from 1-10) for the ADOS.

ADOS	Flo/Ins Evaluative	Shoe Stereotype	Race Evaluative	Race Stereotype	Gender Stereotype
Raw Scores (A+B)	r(29) = .15 [-.21, .49]	r(23) = -.15 [-.48, .39]	r(29) = -.22 [-.59, .46]	r(25) = .07 [-.37, .61]	r(28) = .10 [-.29, .39]
Old Algorithm (A+B)	r(33) = .12 [-.22, .45]	r(25) = -.20 [-.51, .32]	r(33) = -.20 [-.56, .44]	r(27) = .07 [-.37, .56]	r(32) = .04 [-.27, .30]
New Algorithm (SA)	r(29) = .10 [-.27, .45]	r(23) = -.21 [-.55, .37]	r(29) = -.21 [-.59, .47]	r(25) = .08 [-.39, .60]	r(28) = .04 [-.36, .35]
New Algorithm Severity Score (SA)	r(29) = .03 [-.34, .40]	r(23) = -.20 [-.59, .41]	r(29) = -.32 [-.64, .23]	r(25) = .00 [-.44, .41]	r(28) = .01 [-.36, .33]

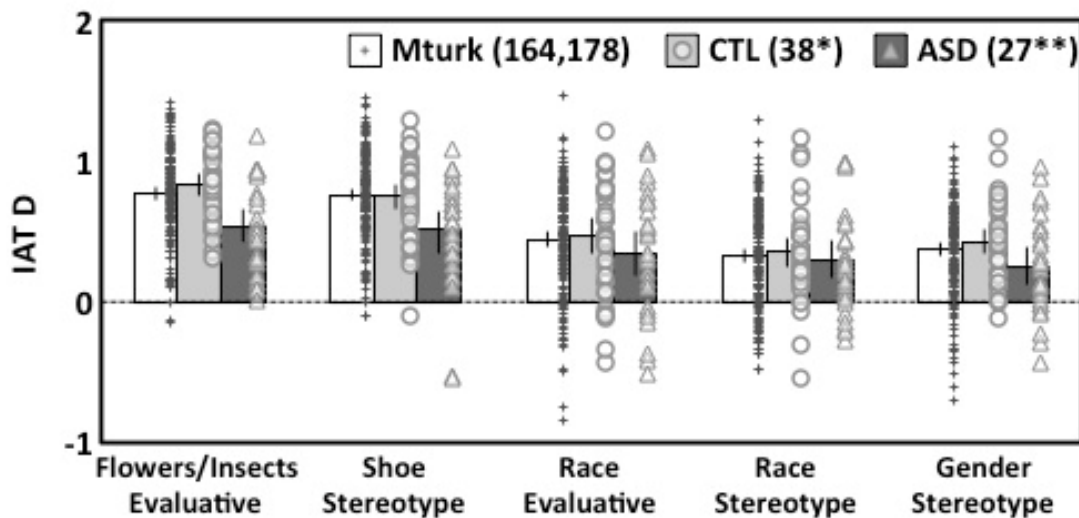
Supplemental Table 8. Internal Consistency Analysis. Correlations between IAT D on practice and test blocks of the IAT.

	ASD	CTL
Flo/ins	0.46 [-0.07, 0.68]	0.22 [-0.09, 0.48]
Shoe Stereo	0.51 [-0.13, 0.80]	0.32 [-0.01, 0.53]
Race Eval	0.61 [0.25, 0.77]	0.47 [0.16, 0.67]
Race Stereo	0.51 [0.17, 0.74]	0.13 [-0.23, 0.48]
Gender Stereo	0.27 [-0.09, 0.56]	0.19 [-0.16, 0.50]

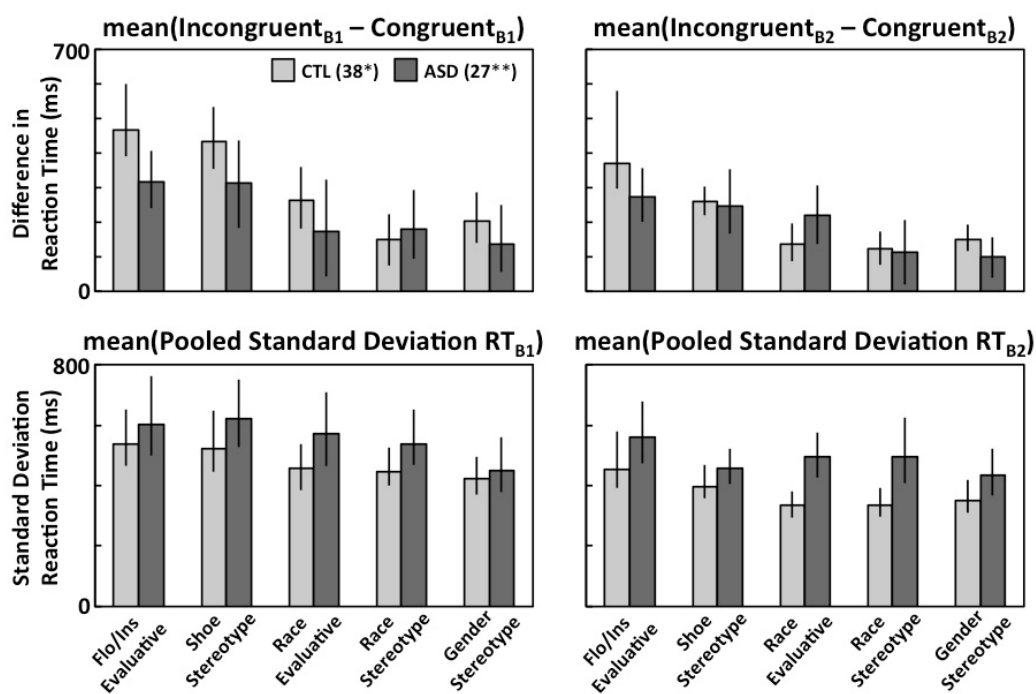
Supplemental Figure 1. Task orders for the various participant groups. In the laboratory, participants (ASD and matched controls) completed either the Eval 1st Order or the Stereo 1st Order (randomly assigned). Online participants were restricted to completing only one of either the Eval or Stereo experiments exclusively. Temporal order proceeds from top to bottom; the order of measures contained within a single box was randomized.

In laboratory participants (ASD/Control)		Online participants (Mturk)	
Eval 1 st Order	Stereo 1 st Order	Eval Order	Stereo Order
Background	Background	Background	Background
RaceEval IAT, FloInsEval IAT, AQ, BAPQ	RaceStereo IAT, GenderStereo IAT, ShoeStereo IAT, AQ, BAPQ	RaceEval IAT, FloInsEval IAT, AQ, BAPQ	RaceStereo IAT, GenderStereo IAT, ShoeStereo IAT, AQ, BAPQ
Semantic Scales (Flowers, Insects)	Semantic Scales (Shoe, Gender Occupation, Race Occupation), MSS	MRS,IMS/EMS, Race Contact Measures, Semantic Scales (Black, White)	MRS,IMS/EMS, Race Contact Measures, MSS, Semantic Scales (Shoe, Gender Occupation, Race Occupation, White, Black)
Break (median=89.6 minutes, min=38 minutes)			
ShoeStereo IAT, RaceStereo IAT, GenderStereo IAT	FloInsEval IAT, RaceEval IAT		
MRS,IMS/EMS, MSS, Race Contact Measures, Semantic Scales (Black, White, Gender Occupation, Race Occupation, Shoe)	MRS,IMS/EMS, Race Contact Measures, Semantic Scales (Black, White)		

Supplemental Figure 2: Means and individual (participant-level) IAT effects for each participant group.



Supplemental Figure 3. Mean RT difference (incongruent – congruent blocks) and RT Variance (standard deviation) for practice (B1) and test (B2) blocks of the IAT. Error bars represent 95% confidence intervals.



Supplemental Figure 4. Mean RT and standard deviations for Congruent and Incongruent block, averaged over practice (B1) and test blocks (B2). Error bars represent 95% confidence intervals.

