Published in final edited form as:
Psychol Sci. 2015 November ; 26(11): 1693–1705. doi:10.1177/0956797615595607.

Implicit Social Biases in People with Autism

Elina Birmingham[1,*], **Damian Stanley**[2,*], **Remya Nair**[2], and **Ralph Adolphs**[2]

[1]Faculty of Education, Simon Fraser University, 8888 University Dr., Burnaby, BC Canada, V5A 1S6

[2]Division of the Humanities and Social Sciences, California Institute of Technology 1200 E. California Blvd., Pasadena, CA, USA, 91125

Abstract

Implicit social biases are ubiquitous and are known to influence social behavior. A core diagnostic criterion of Autism Spectrum Disorder (ASD) is abnormal social behavior. Here we investigated the extent to which individuals with ASD might show a specific attenuation of implicit social biases, using the Implicit Association Test (IAT) across Social (gender, race) and Nonsocial (flowers/insect, shoes) categories. High-functioning adults with ASD showed intact but reduced IAT effects relative to healthy controls. Importantly, we observed no selective attenuation of implicit social (vs. nonsocial) biases in our ASD population. To extend these results, we collected data from a large online sample of the general population, and explored correlations between autistic traits and IAT effects. No associations were found between autistic traits and IAT effects for any of the categories tested in our online sample. Taken together, these results suggest that implicit social biases, as measured by the IAT, are largely intact in ASD.

Keywords

Autism Spectrum Disorder; implicit bias; Implicit Association Test (IAT); stereotype; prejudice

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder diagnosed in part by profound difficulties in social and communicative behaviors, often the most disabling aspect especially in higher functioning individuals (APA, 2013). Despite often normal intellect, people with ASD process faces atypically (see Schultz, 2005), have difficulty interpreting social cues (Baron-Cohen, Wheelwright, & Jolliffe, 1997), and show impaired perspective taking and theory of mind (Moran et al., 2011). In addition, relative to typically developed (TD) individuals, people with ASD attend differently to social stimuli (Dawson et al., 2004), are less responsive to social rewards (Lin, Rangel & Adolphs, 2012), and invest less energy into reputation management (Izuma et al., 2011). Attenuated social motivation has been

Corresponding Author: Elina Birmingham, Mailing Address: Faculty of Education, Simon Fraser University, 8888 University Dr., Burnaby, BC, Canada, V5A 1S6, ebirming@sfu.ca, Phone: 778-782-9392.
*DS and EB contributed equally to this paper

Author Contributions
E.B., D.S., and R.A. developed and designed the study. Testing and data collection were performed by E.B., D.S., and R.N., who also performed the data analysis. E.B., D.S., and R.A. interpreted the results and wrote the paper. All authors approved the final version of the paper for submission.

proposed to catalyze the atypical development of social cognition by reducing exposure to, and therefore fluency with, social stimuli (Chevallier et al., 2012).

Prejudices (the evaluation of a person, place, or thing) and stereotypes (the association between a social group and particular attributes) guide our decision-making in a complex social world. These social biases[1] form early in development: for instance, by age 6, White children explicitly judge White individuals to be preferable to, and more likeable than, Black individuals (Baron & Banaji, 2006), perceive boys to be better at math than girls (Cvencek, Meltzoff, & Greenwald, 2011), and demonstrate explicit knowledge of the stereotypic associations between gender labels and toys (Martin, Wood, & Little, 1990). However, by age 10 children begin to adjust their reporting of such attitudes when explicitly asked about them (Baron & Banaji, 2006), suggesting the emergence of self-presentation concerns and reputation management strategies.

In contrast to explicit reports of bias, implicit measures assess automatic associations between concepts and evaluations/attributes and, as such, are less susceptible to self-presentation effects. For instance, in the Implicit Association Test (IAT), observers must rapidly categorize concepts (e.g., flowers, insects) and evaluations (e.g., good, bad), or concepts (e.g., White person, Black person) and attributes (e.g., intelligent, athletic) using response mappings that are either association-compatible or association-incompatible. The Reaction Time (RT) advantage for association-compatible trials relative to association-incompatible trials indicates the strength of the association between the concepts. The IAT has been used widely to demonstrate the existence of strong implicit biases, even among people who explicitly report not to hold these biases (Baron & Banaji, 2006). Furthermore, performance on the IAT has been shown to predict real-world behaviors despite explicit preferences to act in an unbiased manner (e.g., Stanley et al., 2011; Kubota et al, 2013).

Acquisition of these social biases requires a person to be sensitive to social group information – i.e., to categorize people according to psychologically relevant dimensions [e.g., race, gender, age, sexual orientation; see Bigler & Liben, 2007 for a review]. Furthermore, meaning must be attached to these social groupings: associations between social categories and attributes or evaluations must be encoded as they are observed in a person's environment (Bigler & Liben, 2007). For instance, individuals must pick up on overt statements linking social groups to attributes (e.g., "African Americans are violent"), or covariation occurring in the environment (e.g., all US Presidents have been male). Thus, our key hypothesis, motivated by evidence for impaired processing of social information in ASD as reviewed above (see also Kennedy & Adolphs, 2012), is that social stereotypes and prejudices should be selectively attenuated in individuals with ASD. In contrast, stereotypes and prejudices for nonsocial categories should be unaffected in ASD.

The handful of extant studies on this topic remains inconclusive. Hirschfeld et al. (2007) found that young children with ASD were able to make normal behavioral attributions based on culturally transmitted race and gender stereotypes. This seemingly fluent use of explicit stereotype knowledge in ASD was found despite clear deficits in social cognitive ability

[1]Note: we use the term "bias" to refer to both stereotypes and prejudices

(e.g., theory of mind). In contrast, Kirchner et al. (2012) reported weaker *implicit* social biases in ASD adults: while they showed a significant (nonzero) IAT effect, it was smaller than in controls, suggesting that *implicit* social bias may be attenuated in ASD even when explicit social knowledge is relatively intact. The few studies on social biases in ASD differ greatly in methodology and focused on a narrow range of biases, making comparisons across different domains (e.g., social/nonsocial) or types of bias (evaluative/stereotype) difficult. For instance, studies finding no ASD-linked abnormalities in social bias used measures in which both implicit and explicit processes likely contributed to the behavioral responses (Hirschfeld et al., 2006; da Fonseca et al. 2011), whereas the one study reporting reduced social bias in ASD used only a single IAT (Kirchner et al., 2012). Perhaps most importantly, none of the previous studies included a non-social control condition, severely limiting the conclusions that could be drawn about the social specificity of the findings.

We addressed these shortcomings by systematically investigating the magnitude of social and non-social implicit bias in ASD, as measured by the IAT. In addition, given evidence of distinct neural underpinnings for *evaluative implicit biases* (i.e., concept-evaluation associations) and *stereotype implicit biases* (i.e., concept-attribute associations; see Amodio, 2014 for a review), we included IAT tests for each type of bias to explore possible dissociations in ASD. Experiment 1 tested people with a diagnosis of ASD and matched controls in the laboratory; Experiment 2 extended the investigation to autistic traits in the general population, tested over the Internet.

Method

Participants

For all experiments, we have reported all measures, conditions, data exclusions, and how we determined our sample sizes.

Pilot Study and Sample Size determination—A prior Pilot Study tested 15–22 individuals with ASD (depending on the task) and 38 matched controls. Of these, 14 of the ASD group and 10 of the controls also participated in the main study reported here; the sole criterion for their inclusion here was availability. The Pilot Study found suggestive evidence for reduced implicit social biases in the ASD group (between-groups effect sizes ranged from d=.26 to d=.99 across the multiple IAT tasks used), but suffered from low power due to its small sample size.

Given the initial findings of this pilot study, we recruited participants for the studies described below. Based on the size of our participant pool, as well as ongoing recruitment, and dropout rates, we aimed at a sample of 30 individuals with ASD. Because our ASD population sample size was necessarily limited, we supplemented the in-lab testing of people with ASD with an internet-based study that could provide a larger sample size; based on our prior findings together with the literature we here aimed at an N=200 TD adults as more than sufficient to detect any putative effects.

In Laboratory

ASD: 30 high-functioning adults with ASD were recruited from our participant database. Participants had to be verbal, high functioning (full-scale IQ above 85), located in the greater LA metropolitan area, willing and able to complete all tests, and meet both Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000) and DSM-IV criteria for ASD.

Our sample size and data-collection stopping rule were determined by the size of our participant pool, from which we had access to ~30 individuals, of which 3 were excluded from the final analyses due to extremely slow responding (e.g., unable to complete the task within the time span of their visit to the lab). The 27 remaining ASD participants (23 male; 20 White; 3 Asian; 3 multiracial; 1 Indian) met DSM-IV diagnostic criteria for autism or Asperger syndrome, and met the cutoff scores for ASD on the ADOS (Module 4). Estimates of IQ were assessed using one of the Wechsler tests [Wechsler Abbreviated Scale of Intelligence (WASI) I, II, or III; Wechsler Adult Intelligence Scale (WAIS), III or R; or the Wechsler Intelligence Scale for Children (WISC), III, for one participant whose IQ was assessed during adolescence]. See Table 1 for participant information.

Typically Developed Controls (Controls): 38 neurologically and psychiatrically healthy American adults with similar demographic characteristics to the ASD group and no family history of ASD were recruited through our control database (30 male; 26 White; 5 Asian; 2 Black/African American; 1 American Indian/Alaska Native; 3 multi-racial; 1 "other"/unspecified race). Controls were matched to our ASD sample on age, gender, and IQ (see Table 1). All participants had normal or corrected-to-normal vision and gave informed consent under a protocol approved by the Institutional Review Board of the California Institute of Technology.

Online—A total of 401 adults were recruited to complete either the stereotype (N=201) or evaluative (N=200) IAT tasks using Amazon's Mechanical Turk (MTurk). The participant pool was restricted to MTurk workers located in the United States who had completed >1000 Human Intelligence Tasks (HITs) and had a >98% approval rating. Of this original 401, 59 were excluded for either failing to complete the full battery of tests, for ignoring our instructions not to repeat tasks, for making excessive errors (pressing the incorrect key 2–4 times on a single trial on more than 5% of trials; or pressing the incorrect key more than 4 times on any single trial). Our final sample consisted of 342 American adults (159 females; 287 White; 20 Asian; 15 Black/African American; 3 American Indian/Alaskan Native; 11 multi-racial; 6 "other"/unspecified race). See Table 1.

Materials

Implicit Association Tests—We administered five IATs: three Social [*Gender Stereotype IAT*: male/female+career/family (Nosek, Banaji, & Greenwald, 2002); *Race Stereotype IAT*: white/black+intellectual/athletic (Amodio & Devine, 2006); *Race Evaluation IAT*: white/black+good/bad (Greenwald, McGhee, & Schwartz, 1998)]; and two Nonsocial [*Flower Evaluation IAT*: flowers/insects+good/bad (Greenwald et al., 1998); and a novel *Shoe Stereotype IAT*: sneakers/dress shoes+sports/business - designed specifically to be a non-social comparison for the social stereotype IATs, tapping into concept-attribute rather than concept-evaluation associations]. Detailed descriptions of IAT stimuli can be found in the Supplemental Material.

Autism-related Assessments—The Autism Spectrum Quotient (AQ: Baron-Cohen et al., 2001) and the Broad Autism Phenotype Questionnaire (BAP-Q: Hurley et al., 2007) were administered to all groups. The AQ is widely used to quantify autistic traits in the general population, and often helps to screen for possible ASD (which must then be diagnostically assessed with other instruments). The BAP-Q has been used to quantify autism-like traits both in first-degree relatives of individuals with ASD (Hurley et al., 2007; Sasson, Lam, Childress, Parlier, Daniels, & Piven, 2013) and in the general population (Wainer, Ingersoll, & Hopwood, 2011). Group means of these scores can be found in Table 1.

Explicit scales—The following self-report scales were administered to assess explicit bias. Ratings on each scale were summed and reverse-coded as needed.

Modern Sexism Scale (MSS): 8-item scale that taps into relatively modern attitudes and beliefs about women (Swim, Aikin, Hall, & Hunter, 1995); higher scores reflect more strongly stereotypical beliefs and attitudes about gender.

Modern Racism Scale (MRS): 7-item scale that measures beliefs and attitudes toward Black Americans (McConahay, Hardee, & Batts, 1981); higher scores reflect stronger anti-Black beliefs and attitudes.

Internal vs. External Motivation Scales (IMS/EMS): 10-item measure assessing internal (5 questions) and external (5 questions) motivations for responding in unprejudiced ways (Plant & Devine, 1998); higher scores reflect higher levels of that type of motivation.

Semantic Differential Scales (Greenwald et al., 1998): 7-point scale of the extent to which participants associate Black People, White People, Flowers, and Insects with positive and negative words (e.g. Beautiful=1, Ugly=7). Difference scores were computed (White-Black; Flowers-Insects) such that higher scores reflect pro-flower and pro-White attitudes.

Gender Role Bias: In line with previous studies (M. J. White & White, 2006), to assess gender stereotypes regarding career and family life, we asked participants to rate the extent to which a man or woman would perform better at different tasks (e.g. “take care of the home”; “manage employees”; 1=man; 3=equal; 5=woman). Higher scores reflect more stereotype-congruent beliefs (i.e. Male/Career, Female/Family).

Race Occupation Bias: Participants rated the extent to which a Black person or a White person would be likely to excel at different occupations (e.g. “College Professor”; “Basketball player”; 1=White; 3=equal; 5=Black). Higher scores reflect more stereotype-congruent beliefs (i.e. White/Intellectual, Black/Physical).

Shoe Type Bias: Participants rated the extent to which a dress shoe or a sneaker would be linked to different activities or profession (e.g. “Sports”; “Business”; 1=Sneakers; 3=equal;

5=Dress Shoes). Higher scores reflect more stereotype-congruent beliefs (i.e. Sneakers/Sports, Dress Shoes/Business).

Design and Procedure

Standard procedures for administering the IAT were followed (Lane, Banaji, Nosek, & Greenwald, 2007; Nosek, Greenwald, & Banaji, 2005), which are those recommended by Greenwald, Nosek & Banaji (2003) with the addition of double the practice trials in the reverse single discrimination block to reduce congruent/incongruent block order effects. In addition, both congruent/incongruent block order and response hand were randomly counterbalanced across participants within each of the ASD, CTL, $MTurk_{Eval}$, and $MTurk_{Stereo}$ groups. Explicit scales were completed in Qualtrics and IATs were administered using a modified version of an open source JavaScript web program (by Winter Mason, Steve Allon) based on the standard IAT (Greenwald et al., 1998; Lane et al., 2007). Please see Supplemental Table 2 for additional information on IAT administration.

Laboratory—Participants in the ASD and Control groups received all 5 of the IAT tests, split into two sessions: the Evaluative session (Flower Evaluation IAT, Race Evaluation IAT; order randomized) and the Stereotype session (Gender Stereotype IAT, Race Stereotype IAT, Shoe Stereotype IAT; order randomized). AQ, and BAPQ were administered once, during the first session (see Supplemental Figure 1 for a complete description of test order). The order of these two sessions was counterbalanced across participants and sessions were separated by a median of 89.6 minutes (minimum of 38 min) to reduce practice effects. Total testing duration (excluding the break period between sessions) was approximately 1–1.5 hours.

Online Sample—Participants recruited through MTurk completed only one of the two sessions ($MTurk_{Eval}$ or $MTurk_{Stereo}$) to reduce the possibility of attrition due to testing duration and the necessity for a break separating sessions.

For all participants, explicit scales relevant to the specific session completed (stereotype or evaluative) were always administered after the IATs and the AQ/BAPQ

Data Analysis

Our data analysis focuses on effect size estimates and bootstrapped estimates of 95% confidence intervals (see Cumming, 2014). For the IAT, we report the group means in the form of IAT 'D' scores (an effect size estimate), computed according to the Improved Scoring Algorithm (Lane et al., 2007). Specifically, we (1) deleted trials with RTs greater than 10,000 msec, (2) deleted subjects for whom more than 10% of trials had latencies less than 300 msec, (3) computed the "inclusive" standard deviation for all trials in Practice Blocks (Blocks 3 and 6) and likewise for all trials in Test Blocks (Blocks 4 and 7) (see Supplemental Table 2 for block structure), (4) computed the mean latency for responses for each of Blocks 3, 4, 6 and 7, (5) computed the two mean differences (Incongruent Practice – Congruent Practice) and (Incongruent Test – Congruent Test), (6) divided each difference score by its associated "inclusive" standard deviation, and (7) computed D = the equal-weight average of the two resulting ratios. Data were analyzed such that positive IAT D scores reflect commonly held associations in the United States (flowers/good, insects/bad; dress shoes/business, sneakers/sports, White/good, Black/bad; White/mental, Black/physical; male/career, female/family). Between-groups (CTL vs. ASD) comparisons are represented as Cohen's d effect size estimates. Error bars represent bootstrapped 95% confidence intervals (CIs; mean estimation, random with replacement, 10,000 samples, using Matlab [releases 2013b and 2014a; The MathWorks, Inc., Natick, Massachusetts, United States.]).

Results

IAT Effects

Figure 1a (left panel) shows the IAT D effects for each participant group broken down by individual test (Flo/Ins Eval, Shoe Stereo, Race Eval, Race Stereo, Gender Stereo; plots of individual subject data can be found in Supplemental Figure 2). These data clearly show that IAT effects for the MTurk and CTL samples were very similar, i.e. well within the 95% CIs of each other. In addition, the mean IAT effects observed for our MTurk samples were similar in magnitude to what has been reported in previous studies of the general population: Race Evaluative D=.44, 95% CI [.38, .50] (previous range: .45–.52 [Nosek et al., 2005]), Race Stereotype D=.32; 95% CI [.27, .38] (previous range: .17–.29 [Amodio & Devine, 2006]), and Gender Stereotype D=.37, 95% CI [.32, .41] (previous range: .42–.46 [Nosek et al., 2005]). Although we present data from the whole (heterogeneous) online sample here, in keeping with approaches used by previous studies, in Supplemental Table 3 we present IAT data for the White MTurk participants only (n=287), in addition to presenting data separately for male (n=183) and female (n=159) participants. IAT effects were highly similar across these subsets of participants, with one exception being that female participants showed a stronger Flowers/Insects IAT effect (.85, 95% CI [.78, .91]) than did males (.66, 95% CI [.58, .73]).

Also evident in Figure 1a (left panel) is that overall, the ASD group showed intact IAT effects across all tests (CIs all excluding zero). Furthermore, while IAT effects were slightly reduced relative to the CTL group, there was no evidence of an interaction between diagnosis (CTL vs. ASD) and IAT type (i.e. social vs. nonsocial; or evaluative vs. stereotype). To investigate whether the ASD group showed a selective reduction in social bias, we pooled the IAT D effects into a social mean and a non-social mean (see right-hand panel of Figure 1a), and estimated Cohen's d effect sizes for the difference between group means (CTL-ASD). The effect size estimate for the group difference was large for nonsocial IATs (d=1.03; 95% CI [.57, 1.58]), and medium for social IATs (d=.56, 95% CI [.05, 1.14]), with 95% CIs in each case excluding zero. However, the effect sizes for social and nonsocial IATs did not differ from each other (overlapping 95% CIs). These findings strongly suggest an effect of diagnosis (CTL > ASD), but no interaction between diagnosis and the social content of the IAT[2]. Similarly, the between-groups effect size for Evaluative IATs was medium (Race, Flowers: d=.76, 95% CI [.25, 1.38]), and similar to the effect size for

[2]To ensure that attenuation due to practice did not obscure any findings, we confirmed that a similar pattern emerges when comparing the first non-social and social IAT completed (see Supplemental Table 1).

Stereotype IATs (Gender, Race, Shoe: Stereo d=.75, 95% CI [.25, 1.32]), suggesting that ASD is not associated with a disproportionate impairment in one particular type of bias.

Because our ASD sample size was limited, we supplemented our main analysis by including additional data from our pilot study [non-overlapping participants only: 8 ASD (2 females), 27 CTLs (3 females); see Supplemental Table 4 for demographic information for these groups] for 3 IATs that were common between studies (Gender Stereo, Race Eval, Flowers/Insects Eval). Bringing up our sample to N=100 (35 ASD, 65 CTLS; matched on full scale, verbal, and performance IQ as well as age), we found that the pattern of results did not change: the ASD group had smaller IAT effects than controls, but this was consistent across both social, d=.49 95% CI [.04, 1.02], and nonsocial, d=.70 95% CI [.29, 1.18], tasks. See Supplemental Table 5 for individual IAT D effects for the combined samples.

Explicit Scales

After a final score was calculated for each of the explicit scales (see Methods), we calculated the effect size (Cohen's d) of the group difference (CTL-ASD) for each measure (see Figure 1b), and 95% CIs around these effect sizes. Positive values indicate a larger reported bias in CTL than in ASD (average scores for explicit measures in the online sample can be found in Supplemental Table 6). In contrast to the IAT effects, we find that in the case of explicit bias, there were no tests for which individuals with ASD reported less bias than Controls. In fact, the ASD population expressed stronger race-related (MRS, EMS) and gender-related (MSS) explicit biases than Controls.

Correlations between IAT effects and ADOS (ASD group)

While no group effects were found, it is possible that these summary statistics obscure a relationship between individual differences in autism severity and IAT. To investigate these relationships, we focused on the ADOS (Module 4), the gold standard in the field for quantifying current behavioral autism symptoms. The ADOS consists of a structured interview and interaction between a trained experimenter and the participant (ca. 1 hour), which is videotaped and scored to consensus by trained raters (Lord et al., 2012). New scoring algorithms have recently been developed for the ADOS (see Hus & Lord, 2014), including a new calibrated severity score for a social/affective (SA) domain of the scale. Given this novel scoring algorithm, and some ambiguity around which scoring methods might be most appropriate to use for correlational analyses, we calculated correlations with four ADOS-derived metrics, all highly intercorrelated and all measuring social and emotional behaviors (raw (Communication (A) + Social (B) raw scores), old algorithm (Communication (A) + Social (B) algorithm scores), new algorithm (SA), and new algorithm severity score (SA); Figure 2). This provides both completeness and facilitates comparison with previous and future work. We focus our analysis *a priori* on the newer metrics (new algorithm (SA) and the associated severity score) as these were established to exhibit less covariation with factors unrelated to autism severity (i.e., age and IQ). To maximize sample size, we again included data from our pilot study for the three IATs that overlapped between studies (Gender Stereo, Race Eval, Flower Eval), for a total N=35 in those cells). We found, for the novel scoring methods [New Algorithm (SA) and New Algorithm Severity Score (SA): see third and fourth row of Figure 2], absent-to-weak correlations between ADOS score and IAT D effect (absent-to-weak correlations were also found when considering raw RT differences between congruent and incongruent blocks, See Supplemental Table 7)[3]. Intriguingly, the largest negative correlation was found between the autism severity score and the Race Evaluative IAT ($r(29) = -.39$ [−.64, −.07]), suggesting that those ASD participants who had the most severe autistic symptoms (in terms of social/affective behaviors) also had the lowest implicit race evaluative biases. However, we stress that the reliability of these correlations is limited by our sample size and we note the inconsistency of this finding across the different methods for calculating autism severity.

Correlations between IAT effects and AQ/BAPQ (Control and MTurk Groups)

To supplement the above analysis, we also probed autistic traits in our two other groups without ASD, using the AQ and BAPQ scores. Figure 3 depicts scatter plots of these data and Pearson correlations (r) across all subjects to maximize statistical power. Across all IAT tests, correlation effect sizes were small, with only two instances where 95% CIs excluded zero (Flo/Ins IAT effect was negatively correlated with AQ and BAPQ scores, see Figure 3).

Possible sources of reduced IAT effects in ASD

A final analysis considered possible sources of the global attenuation of IAT D scores that we observed in participants with ASD (Figure 1). The IAT D score is calculated by taking the difference between RTs in the congruent and incongruent conditions and dividing it by the standard deviation of the combined RTs (Greenwald et al., 2003; Lane et al, 2007). Given this, we considered the possibility that increased variability in RTs in individuals with ASD might account for some of the reduced IAT effects in our ASD sample. In two analyses, we (1) looked at raw-RT differences to confirm group-level effects in reaction times (Supplemental Figures 3 and 4); (2) we examined residual IAT D effects after regressing out overall standard deviation in RTs (see Supplemental Results). The first analysis revealed that the pattern of results remained the same when looking at raw-RT differences, ASD participants showed intact social and non-social implicit biases that were generally attenuated but not different (i.e. overlapping 95% CIs) from those of CTL participants. The second analysis revealed that a globally attenuated IAT D effect in ASD persisted even when RT standard deviation was accounted for. Thus, we conclude that increased variability in RTs of ASD participants cannot account for the reduced IAT effects seen in ASD relative to controls.

Finally, to confirm that slowed responding by the ASD group could not account for their reduced IAT effects relative to CTL, we ran similar regressions to that above, but regressing out overall RT, congruent RT, or incongruent RT (3 separate regressions). Again, the overall pattern of results remained the same: Relative to CTL, ASD showed a slightly weaker overall IAT effect in each case (see Supplemental Results).

[3]To ensure that absent correlations were not due to decreased internal consistency in ASD (compared to CTL) participants, we verified the internal consistency of IAT D effects in our two populations (Supplemental Table 8). If anything, internal consistency was higher in our ASD population, ruling out IAT D variability as a factor in weak correlations.

Discussion

Here we present the first systematic investigation of implicit biases in ASD, as measured by the IAT. We examined the extent to which implicit social biases are selectively attenuated in ASD, a disorder characterized by atypical social information processing. We found an overall reduction in the IAT effect in ASD, replicating Kirchner et al.'s (2012) finding of a reduced race evaluative IAT effect in individuals with ASD. However, our study reveals that reductions in implicit bias in ASD are not specific to social categories, but appear to hold across all implicit biases, whether social or nonsocial. Furthermore, as IAT effects were generally intact (i.e., well above zero), we conclude that the basic mechanisms for forming implicit social associations remain intact in ASD.

In addition, based on evidence for a dissociation between valence-based (i.e., evaluations) compared to semantic (i.e., stereotypes) associations (Amodio & Devine, 2006), we explored whether ASD-related abnormalities in these distinct types of bias could be identified. Indeed, there is evidence to suggest that children with Williams Syndrome, a social disorder characterized by hypersociability and lack of social fear, show no evaluative race bias, despite normal gender stereotypes (Santos, Meyer-Lindenberg & Deruelle, 2010). However, our data suggest that both valence-based and semantic associations remain intact in ASD, as group differences did not vary between evaluative (race, flowers) and stereotype (gender, race, shoes) IATs.

We supplemented group-level comparisons with an examination of more continuous relationships across individual differences in autism severity. While correlations between IAT scores and autism severity were generally weak to non-existent, we did find a moderate negative relationship between Race Evaluative IAT effects and ADOS severity scores on social/affective functioning. Although potentially interesting, we hesitate to interpret this result due to the small sample and lack of consistency across different methods for calculating autism-severity (see Figure 2). Instead, we suggest that future work should focus on replicating and further exploring this possible relationship. It may be that a disproportionate effect of autism symptomatology on implicit social evaluative biases is only evident when considering more severe cases of autism (which our sample, by selection, does not have).

Similar to our findings in the lab, we found no relationships between IAT effects and autistic traits in the general population (e.g., online participants), for either social or nonsocial IATs (Figure 2). Given our larger sample size, we suggest that any associations between autistic traits in the general population (as indexed by the BAP-Q and AQ), and implicit associations, as measured by the IAT, are weak at best.

In contrast to our finding of intact-but-reduced IAT effects in ASD, we found no evidence of reduced explicit ratings of bias in ASD. Indeed, if anything, individuals with ASD reported more extreme biases than controls on some measures assessing social biases (Figure 1b). One explanation for this may be that individuals with ASD show less sensitivity to social reputation concerns (Izuma et al, 2011) and are thus less inclined to engage in modulation of their responses on surveys assessing explicit bias. Therefore, at the explicit level our data converge with that of previous studies (Hirschfeld et al., 2007; da Fonseca et al., 2011) in demonstrating that in ASD there is intact knowledge of culturally transmitted stereotypes.

By including social and non-social conditions, our study demonstrated that previously observed attenuation of the IAT effect in ASD (Kirchner et al., 2012) is not specific to social categories, suggesting that it does not result from social impairments present in ASD. Rather, it may result from non-socially-specific cognitive processing differences between ASD and TD individuals. Regression analyses ruled out ‘simple’ explanations that the observed attenuation in IAT D scores was due to increased RT variability or slower responding in the ASD population. Future work should focus on elucidating whether this overall attenuation reflects a true attenuation of implicit biases in a consequential manner, or, is an artifact of task-specific processing demands.

One limitation of the current study is that we used a single measure of implicit bias, the IAT, which requires explicit activation of category-level semantic knowledge. Therefore these data do not address the question of spontaneous activation of implicit biases during the course of everyday social behavior (e.g., when categories are not explicit and salient). Other tasks [e.g. Evaluative priming (Fazio et al., 1995), or the Affect Misattribution Procedure (Payne et al, 2005)] may provide a more direct assay of spontaneous activation of implicit attitudes. We chose here to focus on the IAT for ease of comparison with previous work (e.g. Kirchner et al., 2012) and because it has been well documented to predict social decision-making (e.g., Stanley et al, 2011; Kubota et al, 2013). Furthermore, using only one task allowed us to investigate across multiple theoretically relevant distinctions (e.g. social/nonsocial, stereotype/evaluative) to provide a more nuanced understanding of implicit associations in ASD. While outside the scope of the current study, future work should validate these findings across a range of tasks.

Indeed, it may be that although implicit social biases (at least, as measured by the IAT) remain largely intact in ASD, their downstream integration with other processes leads to atypical social decision-making in the real world. In addition to using other tasks, future research should investigate whether IAT scores in ASD predict deliberative behaviors as they do in TD individuals (e.g., Stanley et al, 2011; Kubota et al, 2013). Combining behavioral studies such as ours with neuroimaging could help to determine whether intact implicit associations in people with ASD are supported by the same neural substrate as in TD individuals. A final important note: because our sample included only high-functioning adults with ASD, future studies that investigate implicit biases across a range of ages and IQ would help to broaden the conclusions that can be drawn from the present study and might well uncover more severe and/or more specific differences than those we report here.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Special thanks to Catherine Armstrong, Tim Armstrong, and Brian Cheng for help with participant recruitment and data collection, to Drs. Daniel Kennedy, Lynn Paul, and Christina Corsello for help in autism assessment, and to the participants and their families. Funded by grants from NSERC to EB (611630), and NIMH (R01MH080721 to RA; K01MH099343 to DS).

References

American Psychiatric Association. DSM 5. American Psychiatric Association; 2013.

Amodio DM. The neuroscience of prejudice and stereotyping. Nature Reviews Neuroscience. 2014; 15:670–682. [PubMed: 25186236]

Amodio DM, Devine PG. Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. Journal of Personality and Social Psychology. 2006; 91(4):652. [PubMed: 17014291]

Baron AS, Banaji MR. The development of implicit attitudes evidence of race evaluations from ages 6 and 10 and adulthood. Psychological Science. 2006; 17(1):53–58. [PubMed: 16371144]

Baron-Cohen S, Wheelwright S, Jolliffe T. Is there a 'language of the eyes'? evidence from normal adults, and adults with autism or asperger syndrome. Visual Cognition. 1997; 4(3):311–331.

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders. 2001; 31(1):5–17. [PubMed: 11439754]

Bigler RS, Liben LS. Developmental intergroup theory explaining and reducing children's social stereotyping and prejudice. Current Directions in Psychological Science. 2007; 16(3):162–166.

Chevallier C, Kohls G, Troiani V, Brodkin ES, Schultz RT. The social motivation theory of autism. Trends in Cognitive Sciences. 2012

Cumming G. The new statistics: Why and how. Psychological Science. 2014; 25(1):7–29. doi: 10.1177/0956797613504966; 10.1177/0956797613504966. [PubMed: 24220629]

Cvencek D, Meltzoff AN, Greenwald AG. Math–gender stereotypes in elementary school children. Child Development. 2011; 82(3):766–779. [PubMed: 21410915]

Da Fonseca D, Santos A, Rosset D, Deruelle C. The “beauty is good” for children with autism spectrum disorders too. Research in Autism Spectrum Disorders. 2011; 5(1):299–304.

Dawson G, Toth K, Abbott R, Osterling J, Munson J, Estes A, Liaw J. Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. Developmental Psychology. 2004; 40(2):271–283. [PubMed: 14979766]

Fazio RH, Jackson JR, Dunton BC, Williams CJ. Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? Journal of personality and social psychology. 1995; 69(6):1013. [PubMed: 8531054]

Greenwald AG, McGhee DE, Schwartz JLK. Measuring individual differences in implicit cognition: The implicit association test. Journal of Personality and Social Psychology. 1998; 74(6):1464. [PubMed: 9654756]

Greenwald AG, Nosek BA, Banaji MR. Understanding and using the implicit association test: I. an improved scoring algorithm. Journal of Personality and Social Psychology. 2003; 85(2):197. [PubMed: 12916565]

Hirschfeld L, Bartmess E, White S, Frith U. Can autistic children predict behavior by social stereotypes? Current Biology. 2007; 17(12):R451–R452. [PubMed: 17580071]

Hurley RS, Losh M, Parlier M, Reznick JS, Piven J. The broad autism phenotype questionnaire. Journal of Autism and Developmental Disorders. 2007; 37(9):1679–1690. [PubMed: 17146701]

Hus V, Lord C. The autism diagnostic observation schedule, module 4: Revised algorithm and standardized severity scores. Journal of Autism and Developmental Disorders. 2014:1–17.

Izuma K, Matsumoto K, Camerer CF, Adolphs R. Insensitivity to social reputation in autism. Proceedings of the National Academy of Sciences. 2011; 108(42):17302–17307.

Kennedy DP, Adolphs R. The social brain in psychiatric and neurological disorders. Trends in Cognitive Sciences. 2012

Kirchner JC, Schmitz F, Dziobek I. Brief report: Stereotypes in autism revisited. Journal of Autism and Developmental Disorders. 2012:1–6. [PubMed: 21360019]

Kubota JT, Li J, Bar-David E, Banaji MR, Phelps EA. The price of racial bias: Intergroup negotiations in the Ultimatum Game. Psychological Science. 2013; 24:2498–2504. [PubMed: 24121413]

Lane, KA.; Banaji, MR.; Nosek, BA.; Greenwald, AG. Understanding and using the implicit association test: IV. what we know (so far) about the method. In: Wittenbrink, B.; Schwarz, N., editors. Implicit measures of attitudes:Progress and controversie. New York: Guilford; 2007. p. 59-102.

Lin A, Rangel A, Adolphs R. Impaired learning of social compared to monetary rewards in autism. Frontiers in Neuroscience. 2012; 6

Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, Rutter M. The autism diagnostic observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders. 2000; 30(3):205–223. [PubMed: 11055457]

Lord, C.; Rutter, M.; Delavore, PC.; Risi, S.; Gotham, K.; Bishop, SL. Autism diagnostic observation schedule, second edition (ADOS-2) manual (part 1): Modulates 1–4. Torrance, CA: Western Psychological Services; 2012.

Martin CL, Wood CH, Little JK. The development of gender stereotype components. Child Development. 1990; 61(6):1891–1904. [PubMed: 2083503]

McConahay JB, Hardee BB, Batts V. Has racism declined in America? It depends on who is asking and what is asked. Journal of Conflict Resolution. 1981; 25(4):563–579.

Moran JM, Young LL, Saxe R, Lee SM, O'Young D, Mavros PL, Gabrieli JD. Impaired theory of mind for moral judgment in high-functioning autism. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(7):2688–2692. doi:10.1073/pnas. 1011734108; 10.1073/pnas.1011734108. [PubMed: 21282628]

Nosek BA, Banaji M, Greenwald AG. Harvesting implicit group attitudes and beliefs from a demonstration web site. Group Dynamics: Theory, Research, and Practice. 2002; 6(1):101.

Nosek BA, Greenwald AG, Banaji MR. Understanding and using the implicit association test: II. method variables and construct validity. Personality & Social Psychology Bulletin. 2005; 31(2): 166–180. [PubMed: 15619590]

Payne BK, Cheng CM, Govorun O, Stewart BD. An inkblot for attitudes: Affective misattribution as implicit measurement. Journal of Personality and Social Psychology. 2005; 89(3):277–293. [PubMed: 16248714]

Plant EA, Devine PG. Internal and external motivation to respond without prejudice. Journal of Personality and Social Psychology. 1998; 75(3):811.

Santos A, Meyer-Lindenberg A, Deruelle C. Absence of racial, but not gender, stereotyping in Williams Syndrome Children. Current Biology. 2010; 20(7):R307–R308. [PubMed: 20392417]

Sasson NJ, Lam KS, Childress D, Parlier M, Daniels JL, Piven J. The broad autism phenotype questionnaire: prevalence and diagnostic classification. Autism Research. 2013; 6(2):134–143. [PubMed: 23427091]

Schultz RT. Developmental deficits in social perception in autism: The role of the amygdala and fusiform face area. International Journal of Developmental Neuroscience. 2005; 23(2):125–141. [PubMed: 15749240]

Stanley DA, Sokol-Hessner P, Banaji MR, Phelps EA. Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proceedings of the National Academy of Sciences. 2011; 108(19):7710–7715.

Swim JK, Aikin KJ, Hall WS, Hunter BA. Sexism and racism: Old-fashioned and modern prejudices. Journal of Personality and Social Psychology. 1995; 68(2):199.

Wainer A, Ingersoll B, Hopwood CJ. The structure and nature of the broader autism phenotype in a non-clinical sample. Journal of Psychopathology and Behavioral Assessment. 2011; 33:459–469.

White MJ, White GB. Implicit and explicit occupational gender stereotypes. Sex Roles. 2006; 55(3): 259–266.

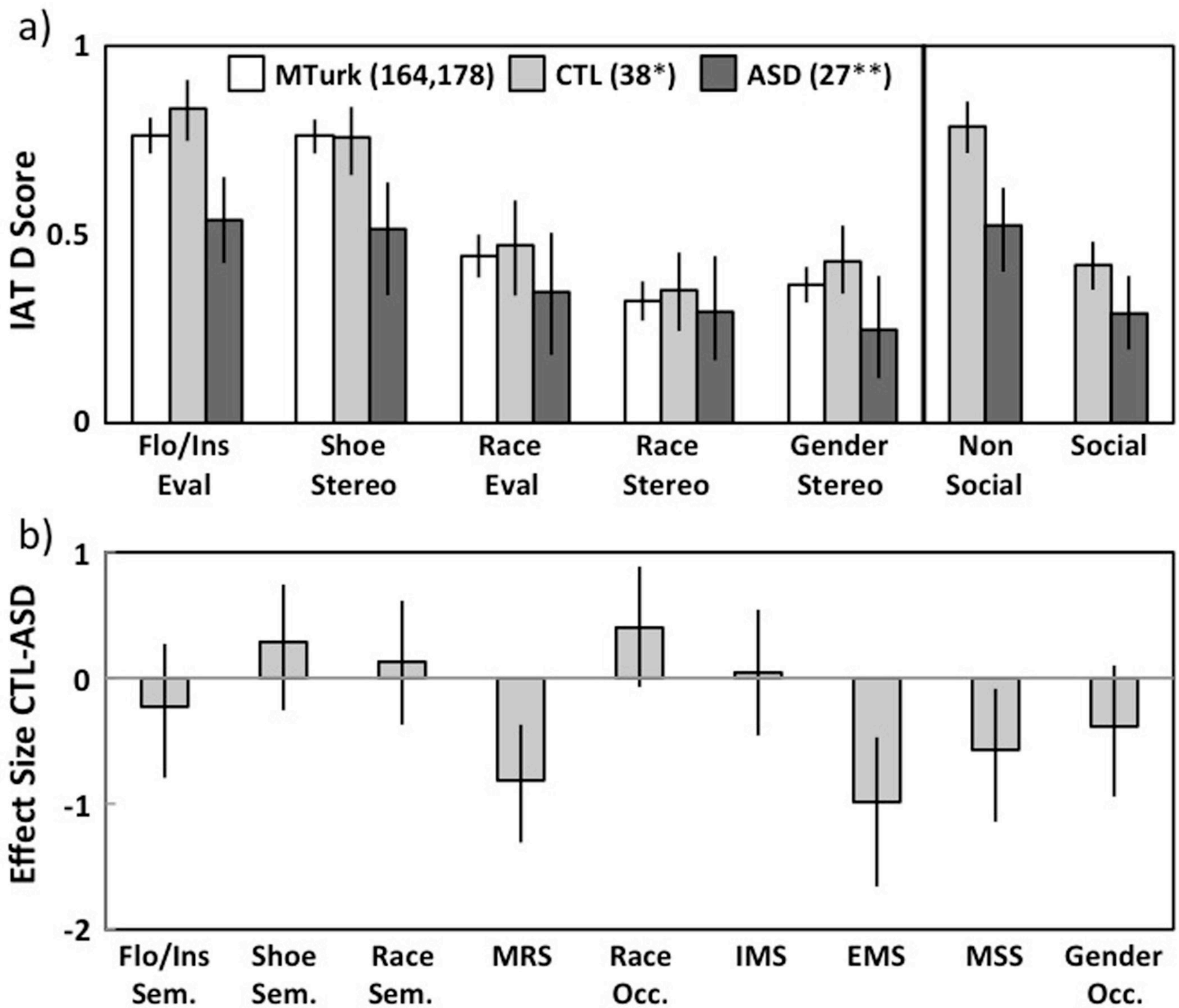

Figure 1.

a) Bar graphs of the average implicit association D scores for each of the five IATs (left panel) completed by the Online (MTurk; white), Controls (CTL; light grey), and Autism Spectrum Disorder (ASD; dark grey) populations and the average D scores across the non-social and social IATs (right panel) for CTL and ASD populations. Note: different online populations completed the Evaluative (N=164) and Stereotype (N=178) tasks (see methods). b) Effect sizes (Cohen's d) of the mean difference between CTL and ASD responses on the explicit measures collected. From left to right: Semantic associations (Flowers/Insects +Good/Bad, Dress Shoes/Sneakers+Sports/Business, Black/White+Good/Bad), Modern Racism Scale (MRS), Black/White+Mental/Physical Occupation (Race Occ.), Internal (IMS) and External (EMS) Motivation Scales, Modern Sexism Scale (MSS), Male/Female +Career/Family Occupation (Gender Occ.). All error bars represent 95% confidence intervals (estimated with a bootstrap procedure). [a]Flo/Ins Eval n=36 CTLs; [b]Gender Stereo n=26 ASD.

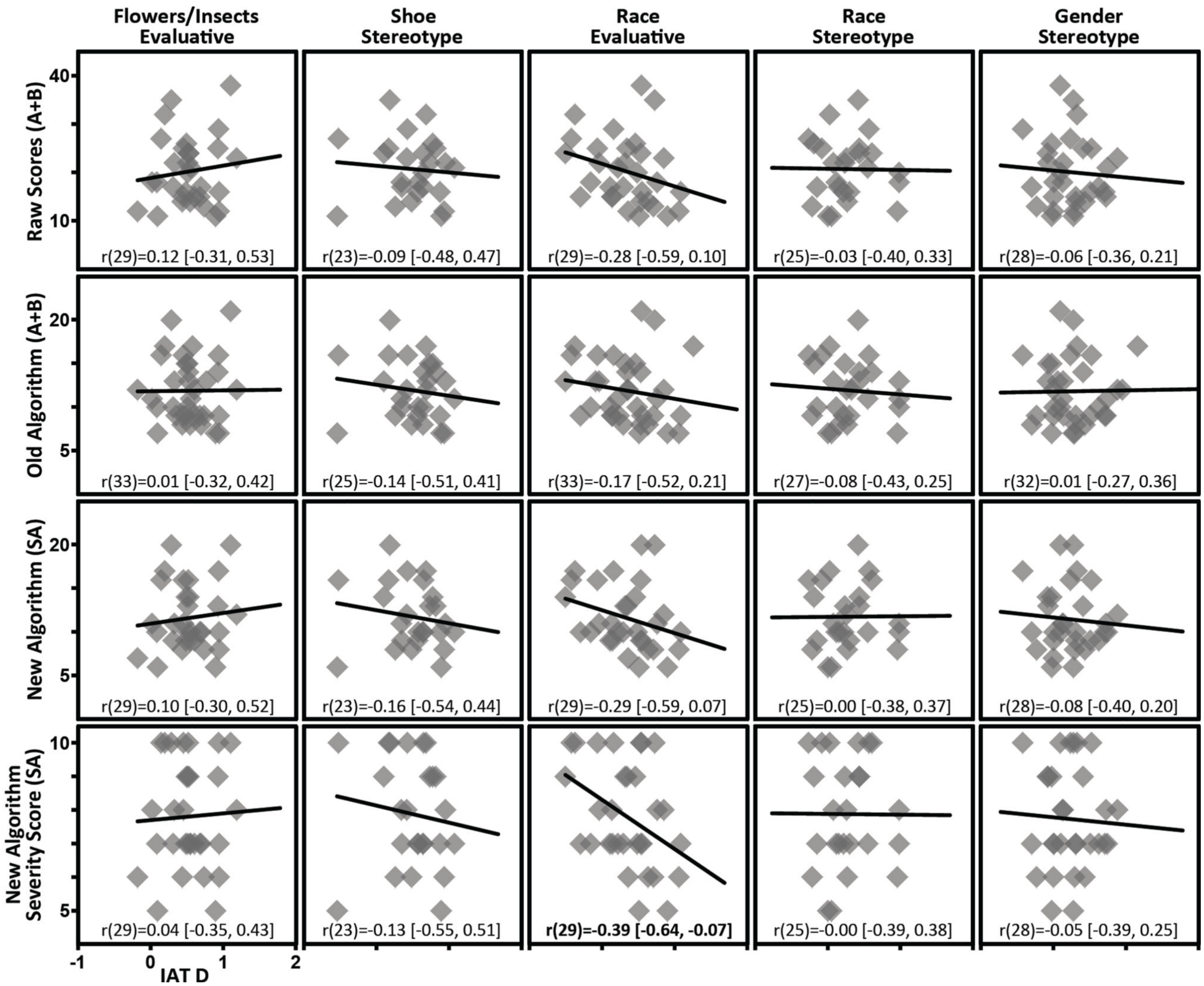

Figure 2.

Correlations (Pearson *r*'s) between IAT D effects and ADOS scores, for the ASD group. Each plot contains Pearson r and bootstrapped 95% CI. Top row: Raw Scores (A+B) = sum of raw scores on the Communication (A) and Social (B) subscales of the ADOS; Second row: Old Algorithm A+B = scores from Communication (A) + Social (B) scoring algorithm; Third row: New Algorithm (SA) = scores from Social Affect algorithm on revised ADOS; and Bottom row: New Algorithm Severity Score (SA) = calibrated Social Affect severity score (scores standardized from 1–10) for the ADOS. Degrees of freedom (in parentheses) for each correlation vary with availability of data points across tasks (no pilot participants took the Shoe IAT; only 2/8 pilot participants took the Race Stereotype IAT), and because 4 participants were missing raw ADOS scores (but had algorithm scores).

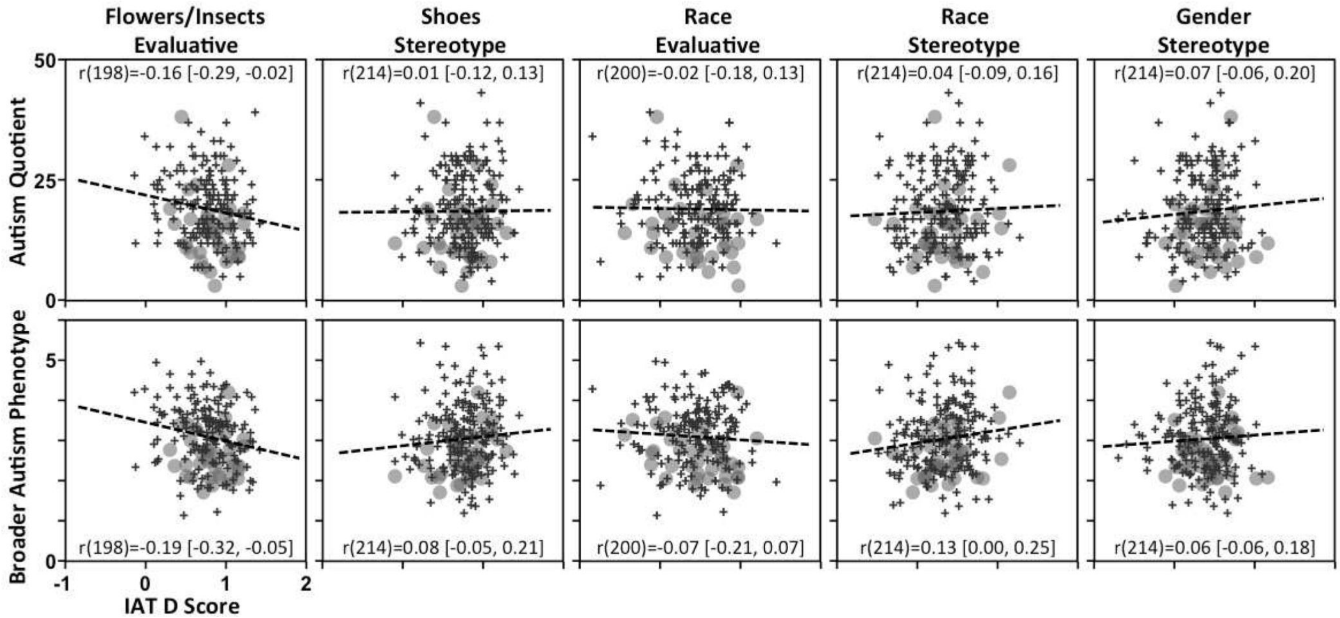

Figure 3.

Top Row: Scatter plots of individual autism-spectrum quotient scores (AQ; y-axis) versus IAT D scores (x-axis) for each of the five IATs completed by the Online (MTurk; crosses) and Controls (CTL; circles) sample. Trend lines and Pearson correlations are for both populations combined. 95% confidence intervals were estimated with a bootstrap procedure. Bottom Row: Similar plots for individual Broad Autism Phenotype Questionnaire scores (BAPQ; y-axis) versus IAT D scores (x-axis) for each of the five IATs completed by the Online (MTurk; crosses) and Controls (CTL; circles) populations. Note: different online populations completed the Evaluative (N=164) and Stereotype (N=178) tasks (see Methods).

Table 1

Demographic information (values represent means ± standard deviations) for each participant group.

	ASD (n=27)	CTL (n=38)	$MTurk_{Stereo}$ (n=178)	$MTurk_{Eval}$ (n=164)
Age (yrs)	30.04 ± 10.78	31.97 ± 9.85	33.40 ± 9.57	34.21 ± 10.58
AQ	28.93 ± 6.48	14.74 ± 6.52	19.35 ± 7.93	19.98 ± 6.99
BAPQ-Total	3.61 ± .56	2.64 ± .56	3.13 ± .85	3.20 ± .77
BAPQ-Aloof	3.50 ± .91	2.73 ± .74	3.37 ± 1.15	3.45 ± 1.09
BAPQ-PragLang	3.53 ± .58	2.44 ± .56	2.63 ± .89	2.69 ± .80
BAPQ-Rigid	3.80 ± .74	2.76 ± .74	3.40 ± .97	3.45 ± .97
FSIQ	107.85 ± 14.76	107.84 + 10.35	--	--
PIQ	108.67 ± 12.94	105.68± 11.18	--	--
VIQ	105.52 ± 17.71	107.97 ± 10.85	--	--
ADOS-A (Cut-off 3/2)	3.70 ± 1.38	--	--	--
ADOS-B (Cut-off 6/4)	8.22 ± 2.36	--	--	--
ADOS-A+B (Cut-off 10/7)	11.93 ± 3.47	--	--	--

ASD: participants with a diagnosis of ASD tested in the laboratory. CTL: controls tested in the laboratory, matched to the ASD group. Participants from the general population were tested on an online version of the stereotype IATs ($MTurk_{Stereo}$) or the evaluative IATs ($MTurk_{Eval}$). ADOS = Autism Diagnostic Observation Schedule. ADOS-A= Communication score; ADOS B=Reciprocal Social Interaction score; ADOS A +B=Communication + Social scores. Higher scores on the ADOS, AQ, and BAPQ indicate greater impairment. For the ADOS, values in parentheses indicate cut-off values for diagnosing Autism/ASD.