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, and [2]Division of the Humanities and Social Sciences, California Institute of Technology

Abstract
Implicit social biases are ubiquitous and are known to influence social behavior. A core diagnostic criterion of autism spectrum disorders (ASD) is abnormal social behavior. We investigated the extent to which individuals with ASD might show a specific attenuation of implicit social biases, using Implicit Association Tests (IATs) involving social (gender, race) and nonsocial (nature, shoes) categories. High-functioning adults with ASD showed intact but reduced IAT effects relative to healthy control participants. We observed no selective attenuation of implicit social (vs. nonsocial) biases in our ASD population. To extend these results, we supplemented our healthy control data with data collected from a large online sample from the general population and explored correlations between autistic traits and IAT effects. We observed no systematic relationship between autistic traits and implicit social biases in our online and control samples. Taken together, these results suggest that implicit social biases, as measured by the IAT, are largely intact in ASD.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder diagnosed in part by profound difficulties in social and communicative behaviors. These difficulties are often the most disabling aspect of the disorder, especially in high-functioning individuals (American Psychiatric Association, 2013). Although people with ASD often have normal intellect, they 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 in reputation management (Izuma, Matsumoto, Camerer, & Adolphs, 2011). It has been proposed that attenuated social motivation in people with ASD catalyzes the atypical development of social cognition by reducing their exposure to, and therefore fluency with, social stimuli (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012).

Prejudices (i.e., stored evaluations of people, places, and things) and stereotypes (i.e., associations between social groups and particular attributes) guide 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), which suggests the emergence of self-presentation concerns and reputation-management strategies.

Corresponding Author:
Damian Stanley, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91101
E-mail: dstanley@caltech.edu

In contrast to explicit measures of bias, implicit measures assess automatic associations between concepts and evaluations or attributes and, therefore, 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., mental, physical), using response mappings that are either congruent or incongruent with typical associations. The different response mappings are used in separate blocks. The reaction time (RT) advantage for congruent trials relative to incongruent trials (*IAT effect*) 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 (Kubota, Li, Bar-David, Banaji, & Phelps, 2013; Stanley, Sokol-Hessner, Banaji, & Phelps, 2011).

Acquisition of these social biases requires a person to be sensitive to social group information—that is, 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 the 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 on covariation occurring in the environment (e.g., all U.S. presidents have been male). Given the evidence for impaired processing of social information in ASD (see also Kennedy & Adolphs, 2012), we hypothesized that social stereotypes and prejudices should be selectively attenuated in individuals with ASD. In contrast, nonsocial stereotypes and prejudices should be unaffected in ASD.

The handful of extant studies on this topic is inconclusive. Hirschfeld, Bartmess, White, and Frith (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 was found despite clear deficits in social-cognitive ability (e.g., theory of mind). In contrast, Kirchner, Schmitz, and Dziobek (2012) reported reduced *implicit* social biases in adults with ASD: Although they showed a significant (nonzero) IAT effect, it was smaller than in control participants. These results suggest 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 have differed greatly in methodology and focused on a narrow range of biases, so that comparisons across different domains (e.g., social vs. nonsocial) or types of bias (evaluative vs. stereotype) are difficult. For instance, studies finding no ASD-linked abnormalities in social bias have used behavioral measures to which both implicit and explicit processes likely contributed (Da Fonseca, Santos, Rosset, & Deruelle, 2011; Hirschfeld et al., 2007), whereas the one study reporting reduced social bias in ASD used only a single IAT (Kirchner et al., 2012). Perhaps most important, none of these previous studies included a nonsocial control condition, which severely limits the conclusions that can be drawn about the social specificity of the findings.

We addressed these shortcomings by systematically investigating the magnitude of social and nonsocial 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 both types of bias to explore possible dissociations in ASD. First, we tested people with a diagnosis of ASD and matched control participants in the laboratory; we then extended the investigation to the general population, testing participants over the Internet. We report all measures, conditions, and data exclusions, as well as how we determined our sample sizes.

Method

Participants

Pilot study and sample-size determination. In a prior pilot study, we tested 15 to 22 individuals with ASD (depending on the task) and 37 matched control participants. Fourteen people in the ASD group and 10 in the control group also participated in the main study reported here; the sole criterion for their inclusion in the main study was availability. The pilot study found suggestive evidence for reduced implicit social biases in the ASD group (*d* for the between-groups effect ranged from 0.26 to 0.99 across the multiple IAT tasks used), but had low power because of the small sample size.

Given the initial findings of this pilot study, we recruited participants for further studies. Given the size of our participant pool, as well as ongoing recruitment and dropout rates, we aimed at a sample of 30 individuals with ASD for in-lab testing. Because our ASD sample size was necessarily limited, we supplemented the in-lab testing with an Internet-based study that could provide a larger sample size; on the basis of our prior findings and results reported in the literature, we decided that a sample size of 200 TD adults would be more than sufficient to detect any putative effects.

Table 1. Background Information for Each Participant Group

Variable	Group			
	ASD ($n = 27$)	Control ($n = 38$)	MTurk: stereotype tasks ($n = 178$)	MTurk: evaluative tasks ($n = 164$)
Age (years)	30.04 (10.78)	31.97 (9.85)	33.40 (9.57)	34.21 (10.58)
AQ score	28.93 (6.48)	14.74 (6.52)	19.35 (7.93)	19.98 (6.99)
BAPQ: total score	3.61 (0.56)	2.64 (0.56)	3.13 (0.85)	3.20 (0.77)
BAPQ: aloof personality	3.50 (0.91)	2.73 (0.74)	3.37 (1.15)	3.45 (1.09)
BAPQ: pragmatic language problems	3.53 (0.58)	2.44 (0.56)	2.63 (0.89)	2.69 (0.80)
BAPQ: rigid personality	3.80 (0.74)	2.76 (0.74)	3.40 (0.97)	3.45 (0.97)
Full Scale IQ	107.85 (14.76)	107.84 (10.35)	—	—
Performance IQ	108.67 (12.94)	105.68 (11.18)	—	—
Verbal IQ	105.52 (17.71)	107.97 (10.85)	—	—
ADOS: Communication (cutoffs: 3/2)	3.70 (1.38)	—	—	—
ADOS: Social Interaction (cutoffs: 6/4)	8.22 (2.36)	—	—	—
ADOS: Communication + Social Interaction (cutoffs: 10/7)	11.93 (3.47)	—	—	—

Note: The table presents means, with standard deviations in parentheses. Participants in the autism spectrum disorder (ASD) group were tested in the laboratory. Participants in the control group were matched to the ASD group and also tested in the laboratory. Participants from the general population (Mechanical Turk, or MTurk, workers) were tested on online versions of the stereotype Implicit Association Tests (IATs) or the evaluative IATs. Autism symptoms were assessed with the Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), the Broad Autism Phenotype Questionnaire (BAPQ; Hurley, Losh, Parlier, Reznick, & Piven, 2007), and the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000); higher scores indicate greater impairment. For the ADOS, the table presents data on the Communication score, the Reciprocal Social Interaction score, and the summed score (Communication + Social Interaction); the cutoffs listed indicate the minimum scores for diagnosing autism and ASD, respectively.

Laboratory samples. Our sample size and data-collection stopping rule were determined by the size of our participant pool. Thirty high-functioning adults with ASD were recruited. To be included in the study, participants had to meet the cutoff scores for ASD on the Autism Diagnostic Observation Schedule (ADOS, Module 4; Lord et al., 2000) and the diagnostic criteria for autism or Asperger's syndrome according to the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994). In addition, they had to be verbal, high functioning (Full Scale IQ above 85), located in the greater Los Angeles metropolitan area, and willing and able to complete all tests. IQ was assessed using one of the Wechsler tests: the first or second edition of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999, 2011); the third or revised edition of the Wechsler Adult Intelligence Scale (Wechsler, 1981, 1997); or the third edition of the Wechsler Intelligence Scale for Children (in the case of 1 participant whose IQ was assessed during adolescence; Wechsler, 1991). Three participants were excluded from the final analyses because of extremely slow responding (e.g., inability to complete the tasks within the time span of their visit to the lab). Table 1 provides further details on the 27 remaining participants (23 male; 20 White, 3 Asian, 3 multiracial, 1 Indian).

Thirty-eight neurologically and psychiatrically healthy American adults with demographic characteristics similar to those of the ASD group and no family history of ASD were recruited through our control database (30 male; 26 White, 5 Asian, 2 Black, 1 American Indian or Alaskan Native, 3 multiracial, 1 unspecified). Control participants were matched to our ASD sample on age, gender, and IQ (see Table 1 for further details).

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 samples. We recruited 401 adults to complete either the stereotype ($n = 201$) or the evaluative ($n = 200$) IAT tasks using Amazon's Mechanical Turk (MTurk). The participant pool was restricted to MTurk workers who were located in the United States, had completed more than 1,000 Human Intelligence Tasks (HITs), and had an approval rating higher than 98%. Of this original 401, 59 were excluded for failing to complete the full battery of tests, for ignoring our instructions not to repeat tasks, or for making excessive errors (pressing the incorrect key two to four times on a single trial on more than 5% of trials or pressing the incorrect key more than four times on any single trial). Our final sample consisted of 342

American adults (183 male; 287 White, 20 Asian, 15 Black, 3 American Indian or Alaskan Native, 11 multiracial, 6 unspecified; see Table 1).

Materials

Implicit Association Tests. We administered five IATs, three social and two nonsocial. The social IATs were the race evaluative IAT, which tested associations between White and Black faces and positive and negative words (Greenwald, McGhee, & Schwartz, 1998); the race stereotype IAT, which tested associations between White and Black faces and words referring to mental and physical attributes (Amodio & Devine, 2006); and the gender stereotype IAT, which tested associations between male and female names and words related to career and family (Nosek, Banaji, & Greenwald, 2002). The nonsocial IATs were the nature evaluative IAT, which tested associations between words referring to flowers and insects and positive and negative words (Greenwald et al., 1998), and a novel shoe stereotype IAT, which tested associations between sneakers and dress shoes and words related to sports and business. The shoe stereotype IAT was designed specifically to be a nonsocial comparison for the social stereotype IATs, tapping into concept-attribute rather than concept-evaluation associations. Descriptions of the IAT stimuli are provided in the Supplemental Material available online.

Autism-related assessments. The Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) and the Broad Autism Phenotype Questionnaire (BAPQ; Hurley, Losh, Parlier, Reznick, & Piven, 2007) were administered to all groups. The AQ is widely used to quantify autistic traits in the general population and to screen for possible ASD (which must then be diagnostically assessed with other instruments). The BAPQ has been used to quantify autism-like traits both in first-degree relatives of individuals with ASD (Hurley et al., 2007; Sasson et al., 2013) and in the general population (Wainer, Ingersoll, & Hopwood, 2011). Table 1 presents mean scores on these instruments for all groups.

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): This eight-item scale assesses 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): This seven-item scale measures beliefs and attitudes toward Black Americans (McConahay, Hardee, & Batts, 1981);

higher scores reflect stronger anti-Black beliefs and attitudes.

- Internal and External Motivation Scales (IMS and EMS): This instrument assesses internal (five questions) and external (five questions) motivations for responding in unprejudiced ways (Plant & Devine, 1998); higher scores reflect higher levels of the indicated type of motivation.
- Semantic differential (Greenwald et al., 1998): On 7-point scales (e.g., *beautiful* = 1, *ugly* = 7), participants rated the extent to which they associated Black people, White people, flowers, and insects with positive and negative words. Difference scores were computed (White − Black; flowers − insects) such that higher scores reflected more stereotype-congruent beliefs (e.g., that flowers are good and insects are bad).
- Gender-role bias: As has been done in previous studies (White & White, 2006), to assess gender stereotypes regarding career and family life, we asked participants to rate the extent to which a man or a woman would perform better at various tasks (e.g., "take care of the home," "manage employees"; 1 = man would perform better, 3 = man and woman would be equal, 5 = woman would perform better). Higher scores reflected more stereotype-congruent beliefs (i.e., that men are better at career-related tasks than women and women are better at family-related tasks than men).
- Race occupation bias: Participants rated the extent to which a Black person or a White person would be more likely to excel at various occupations (e.g., college professor, basketball player; 1 = White more likely to excel, 3 = equal likelihood, 5 = Black more likely to excel). Higher scores reflected more stereotype-congruent beliefs (i.e., that Whites are intellectually inclined and Blacks are physically inclined).
- Shoe-type bias: For this two-item measure, participants rated the relative extent to which a dress shoe and a sneaker would be linked to sports and to business activities (1 = sneakers are more closely related to the activity, 3 = both types of shoes would be appropriate for the activity, 5 = dress shoes are more closely related to the activity). Higher scores reflected more stereotype-congruent beliefs (i.e., that sneakers are associated with sports and dress shoes are associated with business).

Design and procedure

Standard procedures for administering the IAT were followed (Lane, Banaji, Nosek, & Greenwald, 2007; Nosek, Greenwald, & Banaji, 2005). In addition, both the order of congruent and incongruent blocks and the response

hand were randomly counterbalanced across participants within the ASD, control, and two MTurk groups. The explicit scales were completed in Qualtrics, and the IATs were administered using a modified version of an open-source JavaScript Web program (Mason & Allon, 2013) based on the standard IAT (Greenwald et al., 1998; Lane et al., 2007). Table S1 in the Supplemental Material provides additional information on IAT administration, using the nature evaluative IAT as an example.

Participants in the ASD and control groups received all five of the IATs, split into two sessions: the evaluative session (nature evaluative IAT and race evaluative IAT, in random order) and the stereotype session (gender stereotype IAT, race stereotype IAT, and shoe stereotype IAT, in random order). The AQ and BAPQ were administered once, during the first session. The order of these two sessions was counterbalanced across participants, and sessions were separated by a median of 89.6 min (minimum of 38 min) to reduce practice effects. Total testing duration (excluding the break between sessions) was approximately 1 to 1.5 hr.

Participants recruited through MTurk completed only one of the two sessions (i.e., evaluative or stereotype) to reduce the possibility of attrition due to testing duration and due to the necessity of ensuring a break between the sessions. For all participants, explicit scales relevant to the specific session completed (stereotype or evaluative) were always administered after the other measures (see Fig. S1 in the Supplemental Material for a complete outline of test order).

Data analysis

Our data analysis focused on effect-size estimates and bootstrapped estimates of 95% confidence intervals (CIs; see Cumming, 2014). For the IATs, we report group mean D scores (effect-size estimates), computed according to the improved scoring algorithm (Lane et al., 2007). Specifically, we deleted trials with RTs greater than 10,000 ms and subjects for whom more than 10% of trials had RTs less than 300 ms. Next, we computed the inclusive standard deviation of RTs for (a) all trials in the two practice blocks (one incongruent and one congruent block) and (b) all trials in the two test blocks (see Table S1 in the Supplemental Material for block structure). We then computed the mean RT for each of these four blocks, computed the mean RT difference between the incongruent and congruent conditions for practice and test blocks separately, divided each difference score by its associated inclusive standard deviation, and computed D as the equal-weight average of the two resulting ratios. A positive D score indicated that participants held associations that are common in the United States (i.e., flowers–good, insects–bad; dress shoes–business, sneakers–sports; White–good, Black–bad; White–mental, Black–physical; male–career,

female–family). For between-groups (control vs. ASD) comparisons, we report Cohen's d effect-size estimates (Lakens, 2013). We calculated bootstrapped 95% CIs for estimates of the means using bootci.m in MATLAB (releases 2013b and 2014a; The MathWorks, Inc., Natick, MA; 10,000 bootstrap samples per estimate).

Results

IAT effects

Figure 1a shows the mean D score for each IAT for each participant group (plots of individual participants' data are provided in Fig. S2 in the Supplemental Material). These data clearly show that IAT effects for the MTurk and control groups were very similar, that is, well within the 95% CIs of each other. In addition, the mean IAT effects for the MTurk samples were similar in magnitude to what has been reported in previous studies of the general population—race evaluative IAT: $D = 0.44$, 95% CI = [0.38, 0.50] (previous range of reported means: 0.45–0.52; Nosek et al., 2005); race stereotype IAT: $D = 0.32$, 95% CI = [0.27, 0.38] (previous range: 0.17–0.29; Amodio & Devine, 2006); and gender stereotype IAT: $D = 0.37$, 95% CI = [0.32, 0.41] (previous range: 0.42–0.46; Nosek et al., 2005). Figure 1a presents data from the whole (heterogeneous) online sample, but in keeping with approaches used in previous studies, in Table S2 in the Supplemental Material, we have provided IAT data separately for the White MTurk participants ($n = 287$) and 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 effect on the nature evaluative IAT ($D = 0.85$, 95% CI = [0.78, 0.91]) than did males ($D = 0.66$, 95% CI = [0.58, 0.73]).

Figure 1a also indicates that, overall, the ASD group showed intact IAT effects across all tests (i.e., all 95% CIs excluded zero). Furthermore, although the IAT effects of the ASD group were slightly reduced relative to those of the control group, there was no evidence of an interaction between group 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 calculated mean D scores for the social IATs and the nonsocial IATs (see the graph on the right in Fig. 1a) and estimated Cohen's d for the difference between the group means (control − ASD). The effect-size estimate for the group difference was large for the nonsocial IATs ($d = 1.03$, 95% CI = [0.57, 1.58]) and medium for the social IATs ($d = 0.56$, 95% CI = [0.05, 1.14]); in each case, the 95% CI excluded zero. However, the effect sizes for the social and nonsocial IATs did not differ from each other (i.e., the 95% CIs overlapped). These findings strongly suggest an effect of group (i.e., greater IAT effects for the control group than for the ASD group), but no interaction between group and the social content of the IAT.[2] Similarly,
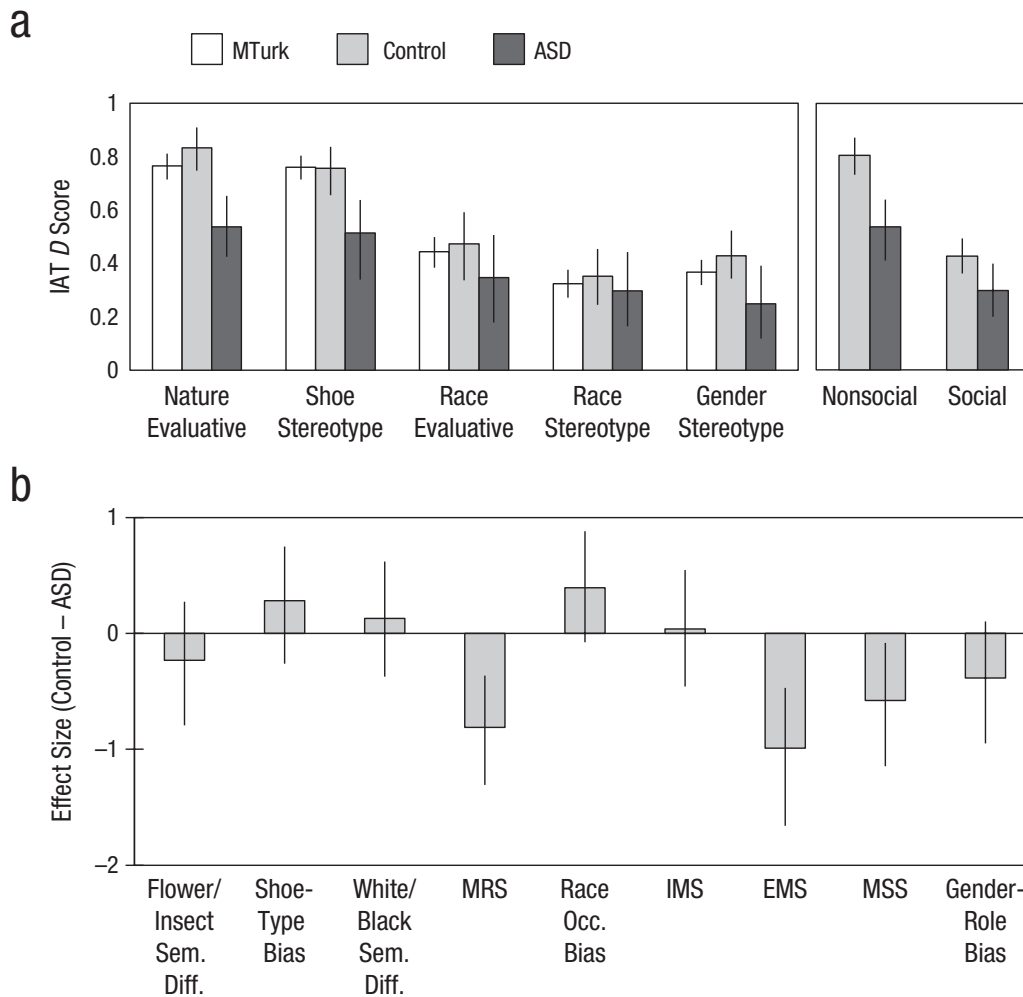
Fig. 1. Comparisons of the participant groups' responses on the Implicit Association Tests (IATs) and the explicit scales. In (a), the mean *D* score is shown for each of the five IATs, separately for the Mechanical Turk (MTurk), control, and autism spectrum disorder (ASD) groups. In addition, the graph on the right shows the average *D* score across the nonsocial IATs and the social IATs, separately for the control and ASD groups. The graph in (b) shows Cohen's *d* for the mean difference between the control and ASD groups' responses on each of the explicit scales (Sem. Diff. = semantic differential; MRS = Modern Racism Scale; Occ. = occupation; IMS = Internal Motivation Scale; EMS = External Motivation Scale; MSS = Modern Sexism Scale). Error bars represent 95% confidence intervals (estimated with a bootstrap procedure). Note that different MTurk groups completed the evaluative (*N* = 164) and stereotype (*N* = 178) IATs (for the control group, *n* = 36 for the nature evaluative IAT and 38 for all other IATs; for the ASD group, *n* = 26 for the gender stereotype IAT and 27 for all other IATs). (For the explicit scales, *n* = 38 for the control group and 27 for the ASD group.)

the between-groups effect size for evaluative IATs (i.e., nature and race evaluative IATs) was medium (*d* = 0.76, 95% CI = [0.25, 1.38]) and similar to the effect size for stereotype IATs (i.e., gender, race, and shoe stereotype IATs; *d* = 0.75, 95% CI = [0.25, 1.32]). These results suggest 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 (nonoverlapping participants only) for the three IATs that were administered in both studies

(gender stereotype, race evaluative, nature evaluative[3]). This increased the size of the ASD group by 8 (2 females) and the size of the control group by 27 (3 females; see Table S4 in the Supplemental Material for demographic information for these participants). Bringing up our sample size to 100 (35 ASD and 65 control participants 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 the control group, but this was consistent across the social tasks (*d* = 0.49, 95% CI = [0.04, 1.02]) and the nonsocial

tasks (d = 0.70, 95% CI = [0.29, 1.18]; see Table S5 in the Supplemental Material for a summary of results for the two groups).

Explicit scales

For each of the explicit scales, we calculated the effect size (Cohen's d; Lakens, 2013) of the group difference (control – ASD) and the 95% CI around the effect size (see Fig. 1b). Positive values would indicate a larger reported bias in the control group than in the ASD group (see Table S6 in the Supplemental Material for the online samples' average scores on the explicit measures). In contrast to the IAT results, the results for the explicit measures revealed no scales on which the ASD group reported less bias overall than the control group (i.e., no positive d values had 95% CIs that excluded zero). In fact, the ASD group expressed stronger race-related (MRS, EMS) and gender-related (MSS) explicit biases than the control group.

Correlations between IAT effects and ADOS scores (ASD group)

Although group effects were consistent across social and nonsocial IATs, it was possible that the summary statistics obscured a relationship between individual differences in autism severity and IAT effects. To investigate this possibility, we focused on the data from the ADOS (Module 4), the gold standard in the field for quantifying current behavioral symptoms of autism. The ADOS consists of a structured interaction between a trained experimenter and the participant (duration of about 1 hr), which is videotaped and scored to consensus by trained raters (Lord et al., 2012). New scoring algorithms, including a new calibrated severity score for a Social Affect (SA) domain, have recently been developed (see Hus & Lord, 2014). Given this novel scoring algorithm, and some ambiguity concerning which scoring methods might be most appropriate to use for correlational analyses, we calculated correlations between IAT effects and four ADOS-derived metrics, all highly intercorrelated and all measuring social and affective behaviors: the total raw score, calculated by adding the Communication and Social Interaction raw scores; the total old-algorithm score, calculated by adding the Communication and Social Interaction algorithm scores; the new-algorithm SA score; and the new-algorithm SA calibrated severity score (SA-CSS). We took this approach both to provide completeness and to facilitate comparisons with previous and future work. We focused our analysis a priori on the newer metrics (i.e., SA and SA-CSS), as these were established to exhibit less covariation with factors unrelated to autism severity (i.e., age

and IQ). To maximize our sample size, we again included data from our pilot study for the three IATs that were included in both studies (gender stereotype, race evaluative, and nature evaluative), for a total n of 35 in those cells.

The results of this correlational analysis are presented in Figure 2. For the two new scoring methods, we found weak to nonexistent correlations between ADOS score and IAT D score (correlations were also weak to nonexistent when we substituted raw RT differences between congruent and incongruent blocks for D scores; see Table S7 in the Supplemental Material).[4] Intriguingly, the largest negative correlation was between SA-CSS and the race evaluative D score, $r(29)$ = −.39, 95% CI = [−.64, −.07], which suggests that those ASD participants who had the most severe autistic symptoms (in terms of social and affective behaviors) also had the lowest implicit evaluative race biases. However, we emphasize that the reliability of these correlations is limited by our sample size, and we note that we did not obtain the same pattern across the different methods for assessing autism symptomatology.

Correlations between IAT effects and AQ and BAPQ scores (control and MTurk groups)

To supplement the analysis just reported, we also looked at correlations between IAT effects and autistic traits in our groups without ASD, using the AQ and BAPQ scores. Figure 3 depicts scatterplots of these data and shows Pearson correlations calculated across all subjects to maximize statistical power. The correlations were small for all the IATs, and there were only two instances in which the 95% CIs excluded zero (the nature evaluative D score was negatively correlated with both AQ and BAPQ scores).

Possible sources of reduced IAT effects in ASD

In a final analysis, we considered possible sources of the global attenuation of IAT D scores that we observed in participants with ASD (Fig. 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, Nosek, & Banaji, 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 (a) looked at raw RT differences to confirm the group-level effects and (b) examined residual IAT D scores after regressing out overall standard deviation in RTs. The first
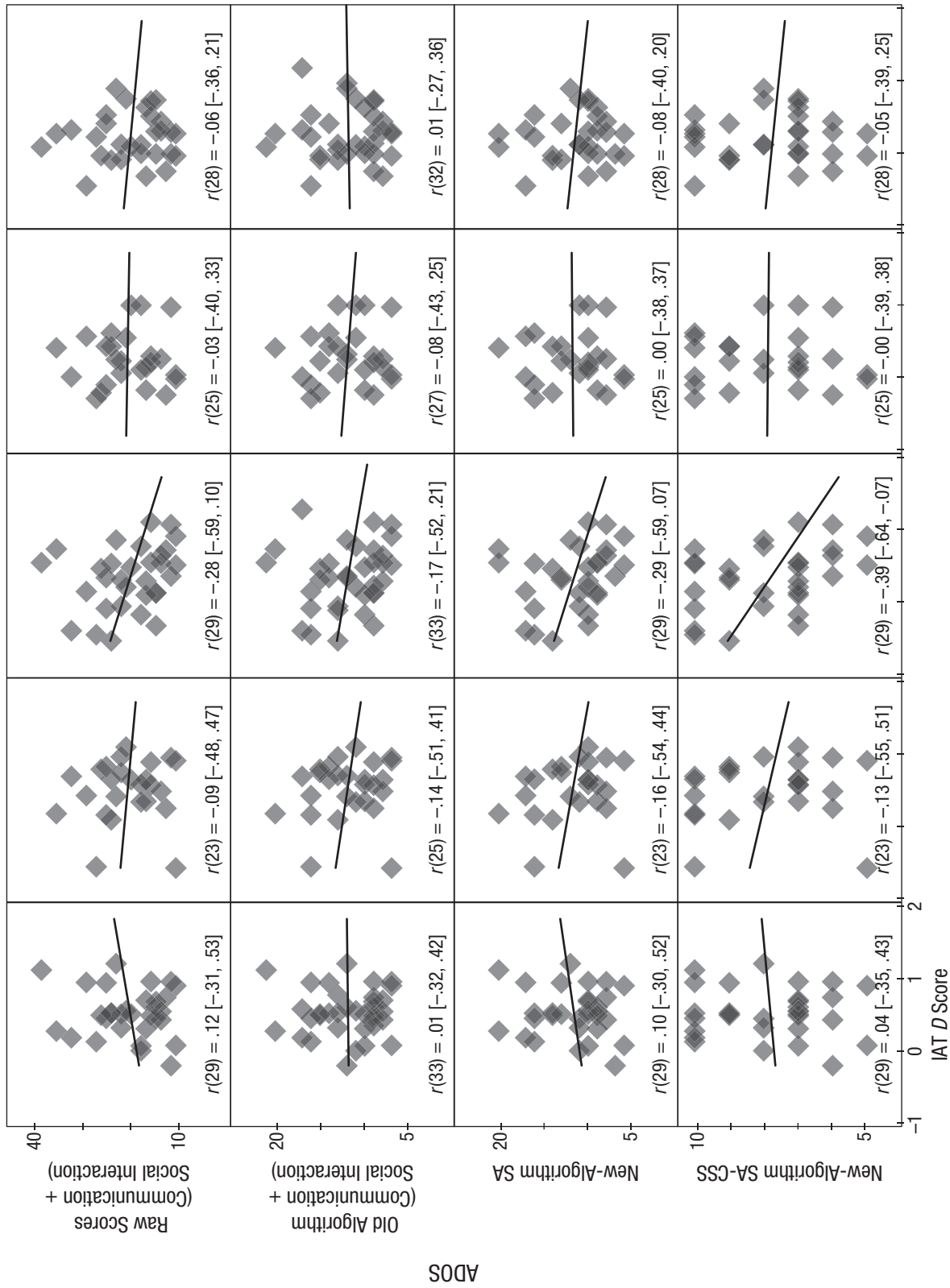
Fig. 2. Results for the autism spectrum disorder group: scatterplots (with best-fitting regression lines) illustrating the correlations between *D* scores on the Implicit Association Tests (IATs) and scores on the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000). Values in square brackets are bootstrapped 95% confidence intervals. Results are shown for four different ADOS metrics: the sum of the raw scores on the Communication and Social Interaction subscales, the sum of the Communication and Social Interaction scores using the old scoring algorithm, the Social Affect (SA) score according to the new algorithm on the revised ADOS, and the calibrated SA severity score (SA-CSS; standardized on a scale from 1 to 10). Note that the degrees of freedom for the correlations vary because the number of data points varied across the IATs (no participants in the pilot study took the shoe stereotype IAT, and only 2 of the participants in that study took the race stereotype IAT, and because raw ADOS scores (but not algorithm scores) were missing for 4 participants.
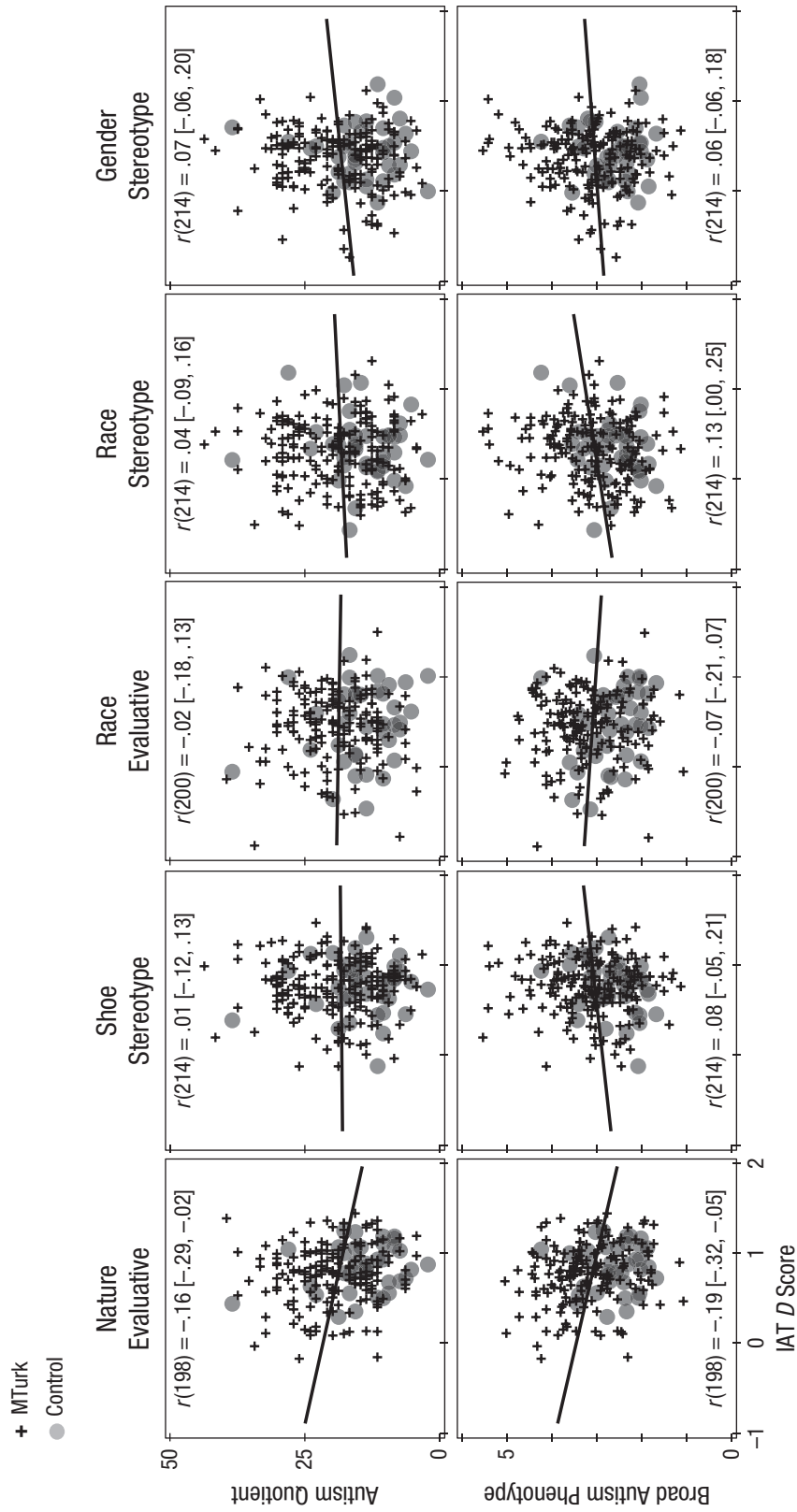
Fig. 3. Results for the control and Mechanical Turk (MTurk) groups: scatterplots (with best-fitting regression lines) illustrating the correlations between D scores on the Implicit Association Tests (IATs) and scores on the Autism Spectrum Quotient (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; top row) and the Broad Autism Phenotype Questionnaire (Hurley, Losh, Parlier, Reznick, & Piven, 2007; bottom row). The regression lines and correlations are for the two groups combined. Values in square brackets are bootstrapped 95% confidence intervals. Note that different online populations completed the evaluative IATs (n = 164) and the stereotype IATs (n = 178).

9

analysis revealed that the pattern of results remained the same when raw RT differences were used instead of *D* scores; ASD participants showed intact social and nonsocial implicit biases that were generally attenuated but not different from those of control participants (i.e., the 95% CIs overlapped; see Figs. S3 and S4 in the Supplemental Material). The second analysis revealed that the globally attenuated IAT effect in ASD persisted even when the standard deviation of RT was accounted for (see Supplemental Results in the Supplemental Material). Thus, we conclude that increased variability in RTs of ASD participants cannot account for the reduced IAT effects they exhibited relative to control participants.

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

Discussion

We have presented 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 for the race evaluative IAT. However, our study revealed 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, given that IAT effects were generally intact in this group (i.e., well above zero), we conclude that the basic mechanisms for forming implicit social associations remain intact in ASD.

In addition, on the basis of evidence for a dissociation between valence-based (i.e., evaluations) and 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 suggesting that children with Williams syndrome, a social disorder characterized by hypersociability and lack of social fear, show no evaluative race bias, despite having 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 the evaluative (race, flower) and stereotype (gender, race, shoe) IATs.

We supplemented group-level comparisons with an examination of more continuous relationships between bias and individual differences in autism severity. Whereas correlations between IAT scores and autism severity were generally weak to nonexistent, we did find a moderate negative relationship between race evaluative IAT *D* score and ADOS severity score for social and affective functioning. Although this finding is potentially interesting, we hesitate to interpret it given the small sample and lack of consistency across different methods for assessing autism symptomatology (see Fig. 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 evident only in severe cases of autism (which our sample, by selection, did not include).

We also found no notable relationships between IAT effects and autistic traits in the general population (online participants and healthy control participants), for either the social or the nonsocial IATs (Fig. 3). Given the larger sample size in this case, we suggest that any associations between autistic traits in the general population (as indexed by the BAPQ and AQ) and implicit associations (as measured by the IAT) are weak at best.

Although we observed reduced (but intact) IAT effects in ASD, we found no evidence of reduced explicit ratings of bias in ASD. Indeed, if anything, participants with ASD reported more extreme biases than control participants on some measures of social biases (Fig. 1b). One explanation for this may be that individuals with ASD are less sensitive 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. Our findings for explicit bias converge with those of previous studies (Da Fonseca et al., 2011; Hirschfeld et al., 2007) in indicating that individuals with ASD have intact knowledge of culturally transmitted stereotypes.

By including social and nonsocial IATs in our study, we demonstrated that the previously observed attenuation of the IAT effect in ASD (Kirchner et al., 2012) is not specific to social categories, which suggests that it does not result from social impairments present in ASD. Rather, it may result from non-socially specific cognitive processing differences between individuals with ASD and TD individuals. Regression analyses ruled out the "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, consequential attenuation of implicit biases (i.e., one that affects actual behavior toward individuals belonging to a social group) 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, our data do not address possible differences in the spontaneous activation of implicit biases during the course of everyday social behavior (e.g., when categories are not explicit and salient). Other procedures (e.g., evaluative priming—Fazio, Jackson, Dunton, & Williams, 1995; the affect misattribution procedure—Payne, Cheng, Govorun, & Stewart, 2005) may provide a more direct assay of spontaneous activation of implicit attitudes. We chose to focus on the IAT because it affords 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., Kubota et al., 2013; Stanley et al., 2011). Further, using only one task allowed us to consider multiple theoretically relevant distinctions (e.g., social vs. nonsocial, stereotype vs. evaluation) to provide a more nuanced understanding of implicit associations in ASD. Although it was outside the scope of the current study to include additional tests of implicit bias, future work should validate our 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 predict deliberative behaviors in individuals with ASD, as they do in TD individuals (e.g., Kubota et al., 2013; Stanley et al., 2011). Combining behavioral studies such as ours with neuroimaging could help to determine whether implicit associations are supported by the same neural substrate in people with ASD as in TD individuals. A final important note is that because our sample included only high-functioning adults with ASD, future studies investigating implicit biases across a range of ages and IQs would help to broaden the conclusions that can be drawn and might well uncover differences more severe or more specific than those we have reported here.

Author Contributions

D. Stanley and E. Birmingham contributed equally to this article. E. Birmingham, D. Stanley, and R. Adolphs developed and designed the study. Testing and data collection were performed by E. Birmingham, D. Stanley, and R. Nair; these three authors also performed the data analysis. E. Birmingham, D. Stanley, and R. Adolphs interpreted the results and wrote the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgments

We thank Catherine Armstrong, Tim Armstrong, and Brian Cheng for their help with recruiting participants and collecting data, and Daniel Kennedy, Lynn Paul, and Christina Corsello for their help in confirming diagnoses of autism. We also wish to thank the participants and their families for committing their time to participate in the research.

Supplemental Material

Additional supporting information can be found at http://pss .sagepub.com/content/by/supplemental-data

Open Practices

All data and materials have been made publicly available via Open Science Framework and can be accessed at https://osf .io/9tu5r/. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/1.%20View%20the%20 Badges/ and http://pss.sagepub.com/content/25/1/3.full.

Notes

1. Note that we use the term *bias* to refer to both stereotypes and prejudices.
2. To ensure that attenuation due to practice did not obscure any findings, we confirmed that a similar pattern emerged when we compared scores from the first nonsocial and social IATs completed (see Table S3 in the Supplemental Material).
3. The race stereotype IAT was introduced into the protocol late, and was administered to only 2 participants from the pilot study. Given that including these two additional data points would not lead to a substantial increase in power, we decided not to include the data for the race stereotype IAT in this analysis.
4. To ensure that the weakness of the correlations was not due to decreased internal consistency in ASD (compared with control) participants, we examined the internal consistency of IAT D scores in our two populations (see Table S8 in the Supplemental Material). If anything, internal consistency was higher in our ASD population, which rules out variability in D scores as a factor in the weak correlations.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*, 670–682.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*, 652–661.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53–58.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, *4*, 311–331. doi:10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. doi:10.1023/A:1005653411471

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, *16*, 162–166.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*, 231–239.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7–29. doi:10.1177/0956797613504966

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, *82*, 766–779.

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

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology*, *40*, 271–283. doi:10.1037/0012-1649.40.2.271

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

Hirschfeld, L., Bartmess, E., White, S., & Frith, U. (2007). Can autistic children predict behavior by social stereotypes? *Current Biology*, *17*, R451–R452.

Hurley, R. S. E., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The Broad Autism Phenotype Questionnaire. *Journal of Autism and Developmental Disorders*, *37*, 1679–1690.

Hus, V., & Lord, C. (2014). The Autism Diagnostic Observation Schedule, Module 4: Revised algorithm and standardized severity scores. *Journal of Autism and Developmental Disorders*, *44*, 1996–2012.

Izuma, K., Matsumoto, K., Camerer, C. F., & Adolphs, R. (2011). Insensitivity to social reputation in autism. *Proceedings of the National Academy of Sciences, USA*, *108*, 17302–17307.

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, *16*, 559–572.

Kirchner, J. C., Schmitz, F., & Dziobek, I. (2012). Brief report: Stereotypes in autism revisited. *Journal of Autism and Developmental Disorders*, *42*, 2246–2251.

Kubota, J. T., Li, J., Bar-David, E., Banaji, M. R., & Phelps, E. A. (2013). The price of racial bias: Intergroup negotiations in the Ultimatum Game. *Psychological Science*, *24*, 2498–2504.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*, Article 863. doi:10.3389/fpsyg.2013.00863

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV. What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 59–102). New York, NY: Guilford Press.

Lin, A., Rangel, A., & Adolphs, R. (2012). Impaired learning of social compared to monetary rewards in autism. *Frontiers in Neuroscience*, *6*, Article 143. Retrieved from http://journal.frontiersin.org/article/10.3389/fnins.2012.00143/full

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., . . . Rutter, M. (2000). 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*, *30*, 205–223.

Lord, C., Rutter, M., Delavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *ADOS-2: Autism Diagnostic Observation Schedule, second edition. Part 1: Modules 1–4*. Torrance, CA: Western Psychological Services.

Martin, C. L., Wood, C. H., & Little, J. K. (1990). The development of gender stereotype components. *Child Development*, *61*, 1891–1904.

Mason, W., & Allon, S. (2013). *Open source, Web-based IAT*. Retrieved from https://github.com/winteram/IAT

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution*, *25*, 563–579.

Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences, USA*, *108*, 2688–2692. doi:10.1073/pnas.1011734108

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration

web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 101–115.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*, 166–180. doi:10.1177/0146167204271418

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affective misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–832.

Santos, A., Meyer-Lindenberg, A., & Deruelle, C. (2010). Absence of racial, but not gender, stereotyping in Williams syndrome children. *Current Biology*, *20*, R307–R308.

Sasson, N. J., Lam, K. S., Childress, D., Parlier, M., Daniels, J. L., & Piven, J. (2013). The Broad Autism Phenotype Questionnaire: Prevalence and diagnostic classification. *Autism Research*, *6*, 134–143.

Schultz, R. T. (2005). Developmental deficits in social perception in autism: The role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience*, *23*, 125–141.

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences, USA*, *108*, 7710–7715.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*, 199–214.

Wainer, A., Ingersoll, B., & Hopwood, C. J. (2011). The structure and nature of the broader autism phenotype in a nonclinical sample. *Journal of Psychopathology and Behavioral Assessment*, *33*, 459–469.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. San Antonio, TX: Psychological Corp.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Psychological Corp.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: Psychological Corp.

Wechsler, D. (1999). *WASI administration and scoring manual*. San Antonio, TX: Psychological Corp.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence—Second Edition*. San Antonio, TX: Psychological Corp.

White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles*, *55*, 259–266.

<u>Supplemental Methods</u>

<u>Stimuli for IATs</u>

Gender Stereotype 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 Evaluative 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 Evaluatiive IAT were paired with words depicting intelligence (mental words: 'Math', 'Brainy', 'Educated', 'Scientist', 'College', 'Genius') and athleticism (physical words: 'Athletic', 'Boxing', 'Basketball', 'Run', 'Jump', 'Football').

Shoe Stereotype 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').

Nature Evaluative 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 (Nature Evaluative, Shoes Stereotype) than for the Social tests (Race Evaluative, Race Stereotype, Gender Stereotype). 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% CI [.57, 1.57]), and medium for social IATs (d=.59, 95% 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.

Table S1. IAT sequence of blocks (example for the Nature Evaluative IAT with congruent condition first, incongruent condition second). Note: These procedures are based on those recommended by Greenwald, Nosek, and Banaji (2003), but the number of practice trials in the reverse single-discrimination block is doubled to reduce effects of the order in which congruent and incongruent blocks are presented.

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

Table S2.   IAT effects and their correlations with Autism Quotient (AQ)/Broad Autism Phenotype Questionnaire (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's 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)
Nature Evaluative	D = .76 [.71, .81]	D = .76 [.70, .81]	D =.66 [.58, .73]	D = .85 [.78, .91]
AQ	r(340) =-.12 [-.28, .04]	r(285) =-.14 [-.30, .04]	r(181) =-.14 [-.37, .11]	r(157) =-.07 [-.26, .15]
BAPQ	r(340) =.-.19 [-.34, -.03]	r(285) =-.18 [-.34, -.002]	r(181) =-.17 [-.41, .10]	r(157) = -.16 [-.36, .06]
Shoes Stereotype	D = .76 [.71, .80]	D = .76 [.71, .80]	D = .78 [.72, .83]	D = .73 [.66, .81]
AQ	r(340) =.01 [-.14, .14]	r(285) =.03 [-.13, .17]	r(181) =.003 [-.20, .18]	r(157) =-.01 [-.21, .18]
BAPQ	r(340) =.05 [-.10, .18]	r(285) =.02 [-.14, .17]	r(181) =.04 [-.16, .22]	r(157) =.03 [-.16, .22]
Race Evaluative	D = .44 [.38, .50]	D = .47 [.40, .53]	D = .49 [.39, .57]	D = .41 [.33, .48]
AQ	r(340) =.04 [-.15, .21]	r(285) =-.001 [-.21, .20]	r(181) =.10 [-.23, .36]	r(157) =-.03 [-.25, .19]
BAPQ	r(340) =-.02 [-.18, .15]	r(285) =-.01 [-.21, .18]	r(181) =.08 [-.19, .36]	r(157) =-.12 [-.32, .09]
Race Stereotype	D = .32 [.27, .38]	D = .32 [.27, .38]	D = .36 [.29, .42]	D =.27 [.19, .35]
AQ	r(340) =.03 [-.11, .18]	r(285) =.03 [-.14, .18]	r(181) =-.07 [-.25, .11]	r(157) =.17 [-.09, .40]
BAPQ	r(340) =.13 [-.01, .26]	r(285) =.11 [-.04, .25]	r(181) =.03 [-.15, .20]	r(157) =.25 [-.001, .43]
Gender Stereotype	D = .37 [.32, .41]	D = .37 [.32, .41]	D = .35 [.29, .41]	D = .39 [.31, .46]
AQ	r(340) =.10 [-.05, .23]	r(285) =.09 [-.07, .24]	r(181) =.17 [-.02, .34]	r(157) =-.01 [-.28, .21]
BAPQ	r(340) =.08 [-.05, .22]	r(285) =.06 [-.08, .21]	r(181) =.16 [-.03, .34]	r(157) =-.02 [-.21, .15]

Table S3. 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 (ES; 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]**	

Table S4. Demographic information (values represent means ± 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. For the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000); higher scores indicate greater impairment. For the ADOS, the table presents data on the Communication score, the Reciprocal Social Interaction score, and the summed score (Communication + Social Interaction); the cutoffs listed indicate the minimum scores for diagnosing autism and ASD, respectively.

	Pilot ASD (n=8)	Pilot CTL (n=27)
Age (yrs)	29.88 ± 10.89	32.96 ± 9.20
Full Scale IQ	106.75 ± 6.20	113.33 + 8.70
Performance IQ	97.88 ± 8.59	110.81 ± 10.51
Verbal IQ	117.38 ± 12.29	112.67 ± 8.42
ADOS Communication (cut-offs 3/2)	4.13 ± 1.73	--
ADOS Social Interaction (cut-offs 6/4)	7.63 ± 3.20	--
ADOS Communication + Social Interaction (cut-offs 10/7)	11.75 ± 4.74	--

Table S5. 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)
Nature Evaluative IAT	0.53 [0.43, 0.63]	0.76 [0.67, 0.83]
Race Evaluative IAT	0.35 [0.20, 0.49]	0.47 [0.38, 0.56]
Gender Stereotype IAT	0.27 [0.16, 0.40]	0.41 [0.34, 0.48]
nonSocial D	**0.53 [0.43, 0.63]**	**0.76 [0.67, 0.83]**
Social D	**0.31 [0.21, 0.42]**	**0.44 [0.38, 0.50]**

Table S6: Explicit scale data (means ± 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/Role = reported associations between Male/Female+Career/Family; Race/Occupation = reported associations between Black/White+Mental/Physical Occupation (Race Occ.)

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

Table S7: Correlations (Pearson's *r*) between ADOS scores and <u>raw RT differences</u> (incongruent minus congruent blocks) on the IAT, for the ASD group (n=25-35 depending on task). Each cell contains Pearson r and bootstrapped 95% CI. Raw Scores (Communication+Social) = sum of raw scores on the Communication and Reciprocal Social Interaction subscales of the ADOS; Old Algorithm (Communication+Social) = scores from Communication + Reciprocal Social Interaction scoring algorithm; New Algorithm (SA) = scores from Social Affect algorithm on revised ADOS; and New Algorithm (SA-CSS) = calibrated Social Affect severity score (scores standardized from 1-10) for the ADOS.

ADOS	Nature Evaluative	Shoe Stereotype	Race Evaluative	Race Stereotype	Gender Stereotype
Raw Scores (Communication+Social)	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 (Communication+Social)	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 (SA-CSS)	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]

Table S8. Internal Consistency Analysis. Correlations (Pearson's r) between IAT D on practice and test blocks of the IAT.

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

Figure S1. 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 1st Order	**Stereo 1st Order**	**Eval Order**	**Stereo Order**
Background	Background	Background	Background
RaceEval IAT, NatureEval IAT, AQ, BAPQ	RaceStereo IAT, GenderStereo IAT, ShoeStereo IAT, AQ, BAPQ	RaceEval IAT, NatureEval 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	NatureEval 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)		

Figure S2: Means and individual (participant-level) IAT effects for each participant group. Note: different online populations completed the Evaluative (N=164) and Stereotype (N=178) tasks (*For Nature Evaluative CTL(N=36); **For Gender Stereotype ASD (N=26); see methods).
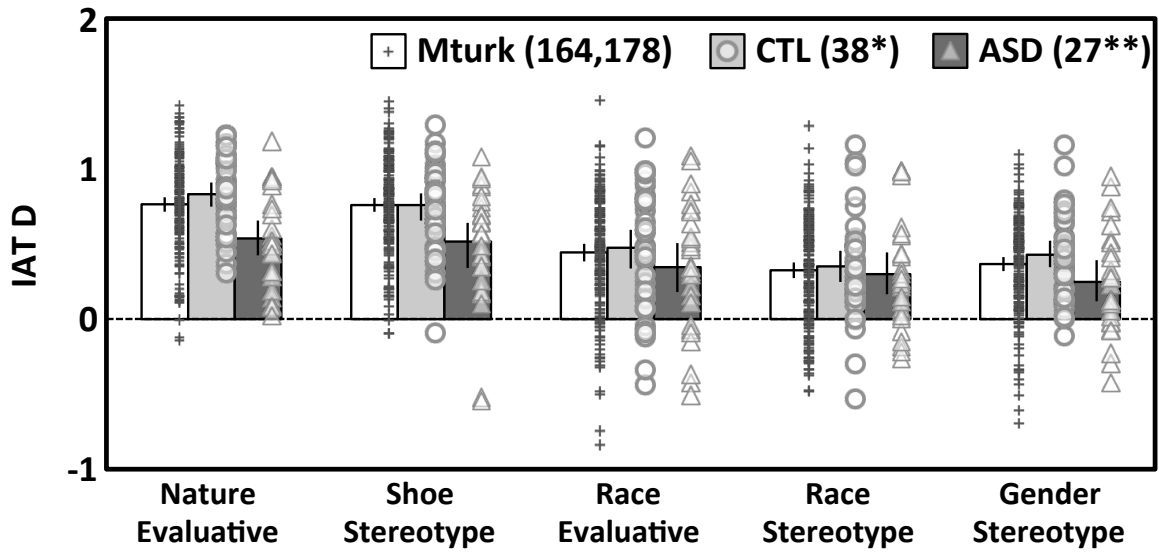
Figure S3. Mean RT difference (incongruent – congruent blocks) and RT Variance (standard deviation) for practice (B1) and test (B2) blocks of the IAT. Error bars represent 95% confidence intervals.
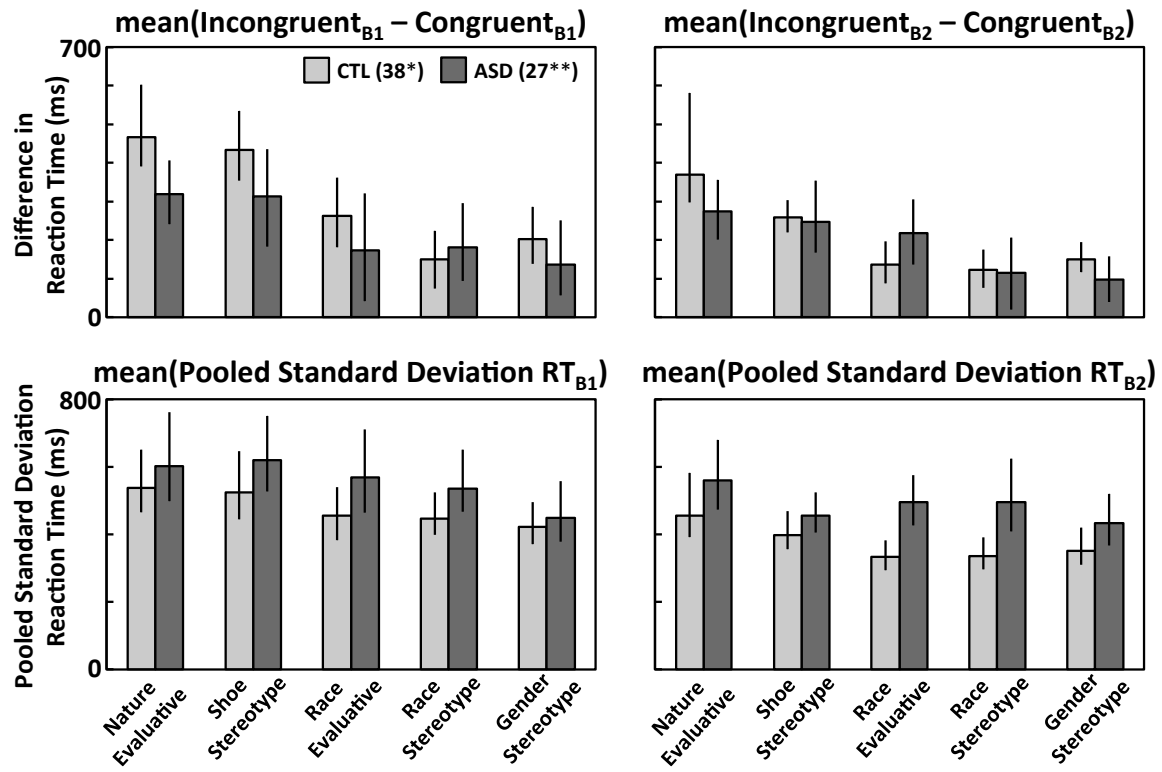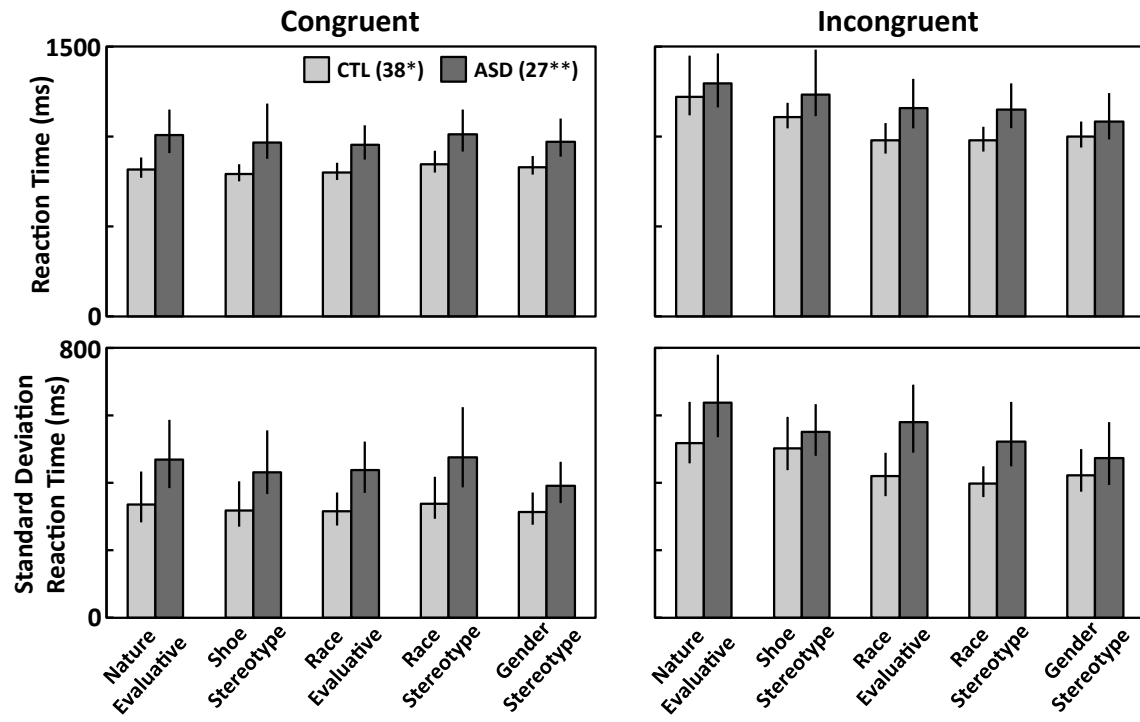
Figure S4. Mean RT and standard deviations for Congruent and Incongruent blocks, averaged over practice (B1) and test blocks (B2). Error bars represent 95% confidence intervals.

Reference

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit

Association Test: I. An improved scoring algorithm. *Journal of Personality and Social*

Psychology, 85, 197–216.