



## Article

<https://doi.org/10.1038/s44220-025-00385-8>

# Phenotypic divergence between individuals with self-reported autistic traits and clinically ascertained autism

Received: 23 April 2024

Accepted: 10 January 2025

Published online: 6 February 2025

Sarah M. Banker , Miles Harrington<sup>3</sup>, Matthew Schafer<sup>4,5,6</sup>, Soojung Na<sup>2,5</sup>, Matthew Heflin , Sarah Barkley<sup>1,3,7</sup>, Jadyn Traywick , Arabella W. Peters , Abigail A. Thinakaran<sup>1,3</sup>, Daniela Schiller , Jennifer H. Foss-Feig<sup>1,3,9</sup> & Xiaosi Gu

Check for updates

While allowing for rapid recruitment of large samples, online research relies heavily on participants' self-reports of neuropsychiatric traits, foregoing the clinical characterizations available in laboratory settings. Autism spectrum disorder (ASD) research is one example for which the clinical validity of such an approach remains elusive. Here we compared 56 adults with ASD recruited in person and evaluated by clinicians to matched samples of adults recruited through an online platform (Prolific; 56 with high autistic traits and 56 with low autistic traits) and evaluated via self-reported surveys. Despite having comparable self-reported autistic traits, the online high-trait group reported significantly more social anxiety and avoidant symptoms than in-person ASD participants. Within the in-person sample, there was no relationship between self-rated and clinician-rated autistic traits, suggesting they may capture different aspects of ASD. The groups also differed in their social tendencies during two decision-making tasks; the in-person ASD group was less perceptive of opportunities for social influence and acted less affiliative toward virtual characters. These findings highlight the need for a differentiation between clinically ascertained and trait-defined samples in autism research.

Online platforms such as Amazon Mechanical Turk and Prolific have become increasingly popular for data collection in human-participant research<sup>1,2</sup>. Through such platforms, researchers can rapidly collect data from hundreds or thousands of participants, allowing for better-powered and more-diverse samples than traditional laboratory data collection. Despite the many benefits of online research, there are also considerable concerns about the quality and validity of such data<sup>3</sup>. For example, previous research has reported low test-retest reliability,

incoherent answers and inattention among online participants on multiple platforms<sup>3–5</sup>. For studies on psychiatric and neurodevelopmental disorders, current online research also relies primarily on self-report surveys to capture traits and diagnoses. Many of these conditions are characterized by altered insight and/or metacognitive awareness<sup>6–8</sup>, and many psychiatric surveys lack diagnostic specificity in the general population<sup>9,10</sup>. Therefore, self-report alone may not be the most accurate way to identify individuals with certain diagnoses or assess observable

<sup>1</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>2</sup>Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>3</sup>Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

<sup>4</sup>The Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>5</sup>Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>6</sup>Department of Psychiatry, Columbia University, New York, NY, USA. <sup>7</sup>Department of Psychology, Stony Brook University, Stony Brook, NY, USA. <sup>8</sup>Department of Psychology, Montclair State University, Montclair, NJ, USA. <sup>9</sup>The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. e-mail: [sarah.banker@nyulangone.org](mailto:sarah.banker@nyulangone.org); [xiaosi.gu@mssm.edu](mailto:xiaosi.gu@mssm.edu)

behavior in certain domains. Without the in-depth clinical characterizations afforded by lab-based research, there is nothing to compare self-report against, making it difficult to determine its generalizability and ecological or clinical validity.

Such concerns may be especially relevant when studying individuals with autism spectrum disorder (ASD), which is often associated with differences in social and self-insight. Previous investigations in ASD have highlighted discrepancies between self- and caregiver-reported autistic traits<sup>11–13</sup> and co-occurring psychiatric conditions<sup>14</sup>. Although one study has identified significant correlations between self- and caregiver-reported autistic trait prominence, the combined measures were more predictive of independent living and employment outcomes than either measure alone, suggesting they do not provide redundant information<sup>15</sup>. In addition, a large-scale study found that self-reported traits among adults evaluated for ASD were not predictive of receiving a diagnosis<sup>16</sup>. Such findings suggest that foregoing the collection of outsider reports, as is often done in online research, may impede the contextualization of self-reported traits and limit the ability to predict clinical diagnoses and objective outcomes. While self-reported traits provide important information about subjective experience and well-being, they may not be an appropriate diagnostic shortcut.

In autistic individuals, discrepancies between self-reported and externally assessed traits may relate to core socioemotional symptoms, including differences in insight and theory of mind (ToM). Differences in ToM—or one's ability to accurately infer the emotions, intentions and mental states of others—may make it difficult for some individuals with ASD to answer questions about others' perceptions of themselves (for example, the prompt from the Broad Autism Phenotype Questionnaire (BAPQ): 'People get frustrated by my unwillingness to bend')<sup>17</sup>. This means that although self-reports may be a useful tool for understanding the internal and lived experiences of autistic individuals, they may not provide a complete, self-aware perspective of the person's behaviors, communication or socioemotional sensitivity and effectiveness as perceived by others or in reference to societal standards.

It is also likely that self-reported traits do not always confer the same meaning in individuals with and without a clinical ASD diagnosis. Although subclinical autistic traits exist in the general population, psychometric investigations into self-report measures have suggested that equal scores in autistic and non-autistic individuals do not necessarily indicate equal levels of autistic traits<sup>18</sup>. In individuals without an ASD diagnosis, a high level of self-reported autistic traits may be more reflective of anxiety surrounding social situations<sup>19,20</sup> and may have little effect on actual social functioning if they are able to compensate through other adaptive behaviors. By contrast, individuals with ASD may underreport their social difficulties due to impaired insight and ToM<sup>11–13</sup>. Thus, interpretations of self-reported traits in the general population may not be directly applicable to ASD.

In this study, we sought to systematically examine trait- and sociocognitive-level similarities and differences between adults with high autistic traits recruited online via self-report and adults with ASD defined via in-person clinical characterization. As social differences are a core feature of ASD, we chose to compare the online and in-person samples on their behavior during dynamic social interaction tasks. Specifically, we aimed to probe strategies for dynamically navigating changing social landscapes and exerting control over social environments, as autistic people show key differences in these areas that can hinder their satisfaction with social relationships and opportunities for employment. We hypothesized that individuals recruited online who self-reported high autistic traits would show distinct social interaction tendencies compared with a clinically defined in-person ASD sample. If true, such results would suggest that self-reported traits alone are not sufficient to identify adults with ASD in online studies.

## Results

### Autistic and other social traits

Participants were enrolled from an online pool consisting of 'unselected' adults from the community (via Prolific,  $n = 502$ ) or had diagnoses of ASD confirmed and enrolled for participation in person (at the Seaver Autism Center in New York City; see Methods for details;  $n = 56$ ). The online sample was further subdivided into 'high-trait' ( $n = 168$ ) and 'low-trait' ( $n = 121$ ) groups based on their total scores on the BAPQ<sup>17</sup>. From within each of these groups, 56 age- and sex/gender-matched participants were selected to match the in-person ASD sample. This resulted in three groups with 56 participants each: high-trait, low-trait and ASD. See Table 1 for demographic characteristics of each group. All comparisons were made on the basis of measurements taken from distinct samples.

As anticipated owing to how the groups were defined, the three groups differed in their self-reported autistic traits, as measured by BAPQ scores ( $F(2,163) = 232.86, P = 1.66 \times 10^{-48}, \eta_{\text{partial}}^2 = 0.74$ ; Fig. 1a). This difference was driven by lower BAPQ scores (indicating fewer traits) in the low-trait group compared with both the high-trait group ( $t(110) = 18.88, P = 2.37 \times 10^{-14}$ , estimated difference = 1.74, 95% confidence interval (CI) [1.52, 1.95], Cohen's  $d = 3.57$ ) and the ASD group ( $t(112) = 18.51, P = 7.26 \times 10^{-14}$ , estimated difference = 1.71, 95% CI [1.49, 1.93], Cohen's  $d = 3.52$ ). The high-trait group and the ASD group did not differ in BAPQ scores ( $t(111) = -0.28, P = 0.957$ , estimated difference = −0.026, 95% CI [−0.25, 0.19], Cohen's  $d = -0.05$ ).

To evaluate the specificity of self-reported traits to ASD versus other disorders characterized by social differences, we also evaluated social anxiety and avoidant personality disorder (AVPD) symptoms. Although considered clinically distinct, social anxiety, AVPD and autism have overlapping features (such as withdrawal from social situations<sup>21–24</sup>) that frequently cluster together in investigations of transdiagnostic factors<sup>25–27</sup>. The groups differed in their social anxiety symptoms ( $F(2,163) = 59.80, P = 3.33 \times 10^{-20}, \eta_{\text{partial}}^2 = 0.42$ ; Fig. 1b), such that the high-trait group had higher scores (indicating more symptoms) than both the low-trait group ( $t(110) = 10.87, P = 5.72 \times 10^{-14}$ , estimated difference = 27.3, 95% CI [21.4, 33.3], Cohen's  $d = 2.06$ ) and the ASD group ( $t(111) = -4.34, P = 9.18 \times 10^{-5}$ , estimated difference = −10.9, 95% CI [−16.9, −4.96], Cohen's  $d = -0.82$ ), and the ASD group had higher scores than the low-trait group ( $t(112) = 6.49, P = 7.33 \times 10^{-9}$ , estimated difference = 16.4, 95% CI [10.4, 22.4], Cohen's  $d = 1.23$ ). Finally, the groups differed in their AVPD traits ( $F(2,163) = 107.84, P = 1.46 \times 10^{-30}, \eta_{\text{partial}}^2 = 0.57$ ; Fig. 1c). The pairwise group differences for AVPD traits follow the same pattern as social anxiety: the high-trait group had higher scores (indicating more symptoms) than both the low-trait group ( $t(110) = 14.58, P = 2.27 \times 10^{-14}$ , estimated difference = 12.50, 95% CI [10.42, 14.58], Cohen's  $d = 2.70$ ) and the ASD group ( $t(111) = -4.18, P = 1.73 \times 10^{-4}$ , estimated difference = −3.66, 95% CI [−5.74, −1.58], Cohen's  $d = -0.79$ ), and the ASD group had higher scores than the low-trait group ( $t(112) = 10.07, P = 0$ , estimated difference = 8.84, 95% CI [6.76, 10.93], Cohen's  $d = 1.91$ ).

In addition to the self-report measures, in-person participants completed the Autism Diagnostic Observation Schedule (ADOS-2; module 4)<sup>28</sup>, considered the 'gold standard' clinical assessment measure for ASD. Surprisingly, there was no significant relationship between self-reported ASD traits measured by BAPQ and those rated by clinicians using ADOS ( $b = 0.025$ , s.e.m. = 0.02,  $t(51) = 1.16, P = 0.251$ , 95% CI [0.0, 1.0],  $\eta_{\text{partial}}^2 = 0.01$ ; Fig. 1d). Broken down by subdomain, there was also no relationship between self- and clinician-rated traits in the restricted and repetitive behavior domain ( $b = 0.12$ , s.e.m. = 0.06,  $t(51) = 1.95, P = 0.057$ , 95% CI [0.0, 1.0],  $\eta_{\text{partial}}^2 = 0.05$ ; Fig. 1e) or the social domain ( $b = 0.05$ , s.e.m. = 0.04,  $t(51) = 1.17, P = 0.249$ , 95% CI [0.0, 1.0],  $\eta_{\text{partial}}^2 = 0.03$ ; Fig. 1f). Such limited agreement between self- and clinician-rated assessments suggests that they may not be measuring the same features of ASD: whereas self-reported assessments can capture subjective internal experiences, clinician-rated assessments

**Table 1 | Group demographic information**

	Clinical ASD (n=56)	High-trait (n=56)	Low-trait (n=56)	Group difference
Age				
Mean (s.d.)	28.07 (8.53)	29.14 (7.49)	29.88 (7.90)	NS
Gender				
% Women (n)	30.3 (17)	41.0 (23)	44.6 (25)	
% Men (n)	51.8 (29)	51.8 (29)	55.45 (31)	P = 0.045
% Non-binary (n)	14.3 (8)	7.1 (4)	0	
% Not reported (n)	3.6 (2)	0	0	
Sex				
% Female (n)	51.8 (29)	46.4 (26)	44.6 (25)	
% Male (n)	48.2 (27)	53.6 (30)	55.45 (31)	NS
% American Indian or Alaska Native (n)	0 (0)	0 (0)	1.8 (1)	
% Asian (n)	10.7 (6)	10.7 (6)	14.3 (8)	
% Black (n)	19.7 (11)	17.9 (10)	7.1 (4)	NS
% White (n)	57.1 (32)	60.7 (34)	67.9 (38)	
% Other (n)	12.5 (7)	10.7 (6)	8.9 (5)	
IQ				
Mean (s.d.)	112.38 (16.26)	—	—	—
Cognitive ability				
Mean (s.d.)	—	7.40 (3.42)	7.02 (3.53)	NS
Ethnicity				
% Hispanic/Latino (n)	23.2 (13)	8.9 (5)	10.7 (6)	NS
Employment				
% Employed (n)	48.2 (27)	58.9 (33)	69.6 (39)	
% Unemployed (n)	51.8 (29)	33.9 (19)	26.8 (15)	P = 0.033
% Not reported (n)		7.2 (4)	3.6 (2)	
Household income				
% 10–50k (n)	42.9 (24)	58.9 (33)	19.6 (11)	
% 50–100k (n)	19.6 (11)	30.4 (17)	53.6 (30)	P = 9.01×10 <sup>-6</sup>
% >100k (n)	37.5 (21)	10.7 (6)	26.8 (15)	
Education				
% Graduate school (n)	17.9 (10)	5.4 (3)	14.3 (8)	
% College (n)	25.0 (14)	32.1 (18)	30.4 (17)	
% Some college (n)	25.0 (14)	37.5 (21)	41.1 (23)	
% High school (n)	23.2 (13)	21.4 (12)	14.3 (8)	NS
% Some high school (n)	7.1 (4)	3.6 (2)	0 (0)	
% No high school (n)	1.8 (1)	0 (0)	0 (0)	

Demographic information is shown for the ASD, high-trait and low-trait groups. Kruskal Wallis tests (nonparametric) examined differences between continuous variables as they were not normally distributed. Chi-squared tests examined differences between factor variables, with Monte Carlo simulation ( $B=10,000$ ) in cases with small frequency counts ( $n<5$ ). Cognitive ability in the online sample was measured using the 16-item International Cognitive Ability Resource test<sup>59</sup>. NS, not significant.

may capture external presentation of traits. Our results suggest that, in ASD, these two domains do not always agree.

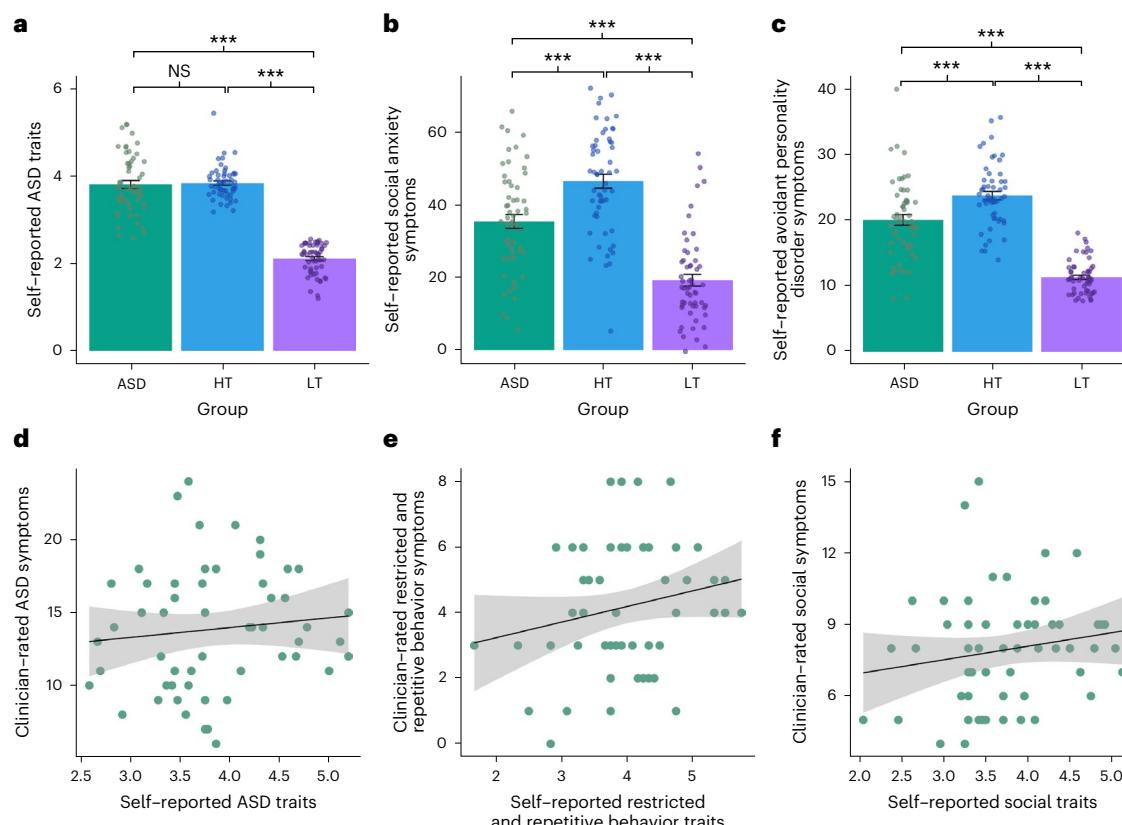
### Social behavior

As social differences are a hallmark feature of ASD, we chose to compare our groups on their behavior in two dynamic social interactions tasks. The paradigms outlined in the following allow for the quantification of complex social processes, including exertion of social control and navigation through ‘social space’.

**Social controllability.** Social controllability, or one’s ability to influence other people, is crucial for achieving optimal behavior during

dynamic interactions and, subsequently, for mental wellbeing. To influence someone, one must first have a sense of that individual’s current and future thoughts, feelings and objectives. We hypothesized that impairments in ToM and social prediction may make it more difficult for individuals with ASD to exert social influence<sup>29–33</sup>.

To measure social controllability, we used a monetary exchange task<sup>26,27,34,35</sup> modified from the ultimatum game, in which participants decide whether to accept or reject proposed splits of US\$20 offered by players from two independent teams (Fig. 2a; see Methods for details). Unbeknown to participants and different from the traditional ultimatum game, participants have control over the offers proposed by one of the teams ('controllable condition'). Specifically, participants can



**Fig. 1 | Trait comparisons.** **a**, The ASD ( $n = 56$  participants) and high-trait (HT;  $n = 56$  participants) groups had comparable levels of self-reported autistic traits (measured via BAPQ; two-sided pairwise comparisons using estimated marginal means, with confidence intervals and  $P$  values adjusted for multiple comparisons using the Tukey method:  $t(111) = -0.28$ ,  $P = 0.957$ , estimated difference =  $-0.026$ , 95% CI  $[-0.25, 0.19]$ , Cohen's  $d = -0.05$ ; mean ASD: 3.82, mean HT: 3.85, mean low-trait (LT): 2.11). **b,c**, Investigation into traits of other disorders characterized by social impairment revealed that, compared with both other groups ( $n = 56$  participants each), the high-trait group ( $n = 56$  participants) self-reported a higher level of social anxiety (two-sided mixed-effects model with random intercept for matched pair ID:  $F(2,163) = 59.80$ ,  $P = 3.33 \times 10^{-20}$ ,  $\eta_{\text{partial}}^2 = 0.42$ ; mean ASD: 35.39, mean HT: 46.43, mean LT: 19.21 (b)) and avoidant personality disorder (AVPD) symptoms (two-sided mixed-effects model with random intercept for

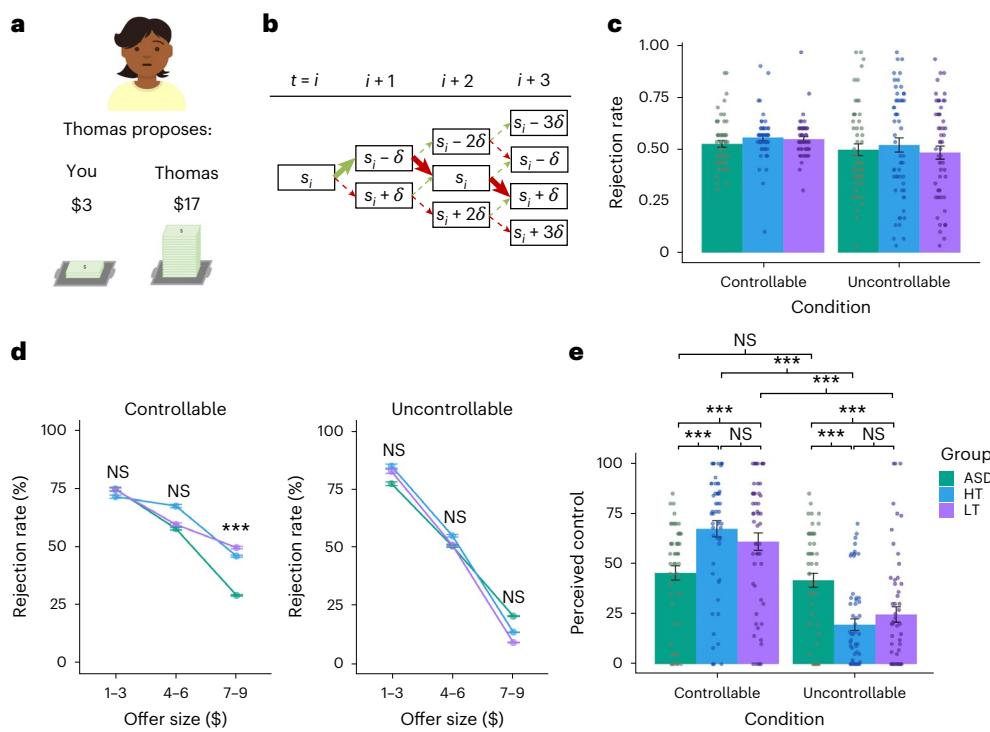
matched pair ID:  $F(2,163) = 107.84$ ,  $P = 1.46 \times 10^{-30}$ ,  $\eta_{\text{partial}}^2 = 0.57$ ; mean ASD: 20.09, mean HT: 23.80, mean LT: 11.36 (c)). **d**, In the in-person ASD group ( $n = 56$  participants), there was no relationship between clinician-rated autistic traits measured via ADOS (mean = 13.85) and self-reported autistic traits measured via BAPQ (two-sided general linear model:  $b = 0.025$ , s.e.m. = 0.02,  $t(51) = 1.16$ ,  $P = 0.251$ , 95% CI  $[-0.018, 0.067]$ ,  $\eta_{\text{partial}}^2 = 0.01$ ). **e,f**, Broken down by subscales, there was no agreement in the restricted and repetitive behavior domain (RRB; general linear model:  $b = 0.12$ , s.e.m. = 0.06,  $t(51) = 1.95$ ,  $P = 0.057$ , 95% CI  $[0.0, 1.0]$ ,  $\eta_{\text{partial}}^2 = 0.05$  (e)) or the social domain (general linear model:  $b = 0.05$ , s.e.m. = 0.04,  $t(51) = 1.17$ ,  $P = 0.249$ , 95% CI  $[0.0, 1.0]$ ,  $\eta_{\text{partial}}^2 = 0.03$  (f)). Error bars represent the s.e.m.; gray shading around regression lines represents the 95% confidence interval. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

increase future offers by rejecting current ones or decrease future offers by accepting the current ones. At the end of the task, participants rate how much control they believed they had over players from each team (Methods).

Because participants can raise offers only by rejecting a current proposal, we first sought to characterize their rejection rate both overall and as a function of offer size. We found that the three groups showed similar overall rejection rates during the task ( $F(2,281) = 0.77$ ,  $P = 0.46$ ,  $\eta_{\text{partial}}^2 = 0.006$ ; Fig. 2c). Breaking rejection rate down by offer size, we found that, while the groups showed similar rejection rates for low offers ( $F(2,74) = 0.20$ ,  $P = 0.82$ ,  $\eta_{\text{partial}}^2 = 0.005$ ) and medium offers ( $F(2,162) = 1.67$ ,  $P = 0.29$ ,  $\eta_{\text{partial}}^2 = 0.02$ ), the ASD group rejected a smaller percentage of high offers ( $F(2,122) = 6.12$ ,  $P = 0.009$ ,  $\eta_{\text{partial}}^2 = 0.09$ ; Fig. 2d) compared with both low-trait ( $t(92) = -3.34$ ,  $P = 0.005$ , estimated difference =  $-0.20$ , 95% CI  $[-0.36, -0.05]$ , Cohen's  $d = -0.72$ ) and high-trait ( $t(86) = -2.67$ ,  $P = 0.025$ , estimated difference =  $-0.16$ , 95% CI  $[-0.30, -0.02]$ , Cohen's  $d = -0.57$ ) online groups. Patterns of rejection rates did not differ across groups for the uncontrollable condition (low:  $F(2,111) = 1.19$ ,  $P = 0.46$ ,  $\eta_{\text{partial}}^2 = 0.02$ ; medium:  $F(2,111) = 0.46$ ,  $P = 0.71$ ,  $\eta_{\text{partial}}^2 = 0.006$ ; high:  $F(2,111) = 3.82$ ,  $P = 0.075$ ,  $\eta_{\text{partial}}^2 = 0.06$ ), with each group showing the highest rejection rates for low offers (\$1–\$3) and the

lowest rejection rates for high offers (\$7–\$9). Together, these results suggest that high-trait online participants behaved more similarly to the low-trait online group than to the clinical ASD group during controllable social interactions, whereas the clinical ASD group demonstrated distinctly reduced ability to exert control.

We next investigated whether participants differed in their subjective perception of the controllability that they had. Indeed, we detected a significant group-by-condition interaction on perceived control ratings ( $F(2,269) = 18.52$ ,  $P = 2.91 \times 10^{-8}$ ,  $\eta_{\text{partial}}^2 = 0.12$ ; Fig. 2e). In the controllable condition, the ASD group perceived less control than both the high-trait group ( $t(110) = -3.79$ ,  $P = 0.0007$ , estimated difference =  $-21.6$ , 95% CI  $[-35.2, -8.1]$ , Cohen's  $d = -0.73$ ) and the low-trait group ( $t(111) = -2.64$ ,  $P = 0.025$ , estimated difference =  $-15.1$ , 95% CI  $[-28.7, -1.5]$ , Cohen's  $d = -0.51$ ); the high- and low-trait groups did not differ from each other ( $t(107) = 1.16$ ,  $P = 0.48$ , estimated difference =  $6.5$ , 95% CI  $[-6.8, 19.8]$ , Cohen's  $d = -0.22$ ). In the uncontrollable condition, the ASD group reported having more control than both the high-trait group ( $t(110) = 4.46$ ,  $P = 5.93 \times 10^{-5}$ , estimated difference =  $21.8$ , 95% CI  $[10.2, 33.5]$ , Cohen's  $d = -0.86$ ) and the low-trait group ( $t(111) = 3.38$ ,  $P = 0.003$ , estimated difference =  $16.6$ , 95% CI  $[4.9, 28.3]$ , Cohen's  $d = 0.65$ ); the high- and low-trait groups once again



**Fig. 2 | Social controllability.** **a**, As shown in the representative task screen, the social control task involved participants accepting or rejecting splits of \$20 proposed by members of two virtual teams. **b**, Participants played the game with two different teams sequentially, the order of which was counterbalanced. With one of the teams (controllable condition), participants could increase future offers by rejecting the current one, or decrease future offers by accepting the current one. **c**, All groups ( $n = 56$  participants each) showed comparable overall rejection rates for both conditions (two-sided mixed-effects model with random intercept for matched pair ID:  $F(2,281) = 0.77, P = 0.46, \eta_{\text{partial}}^2 = 0.006$ ; mean ASD controllable: 52.4%, mean HT controllable: 55.5%, mean LT controllable: 54.7%, mean ASD uncontrollable: 49.6%, mean HT uncontrollable: 51.9%, mean LT uncontrollable: 48.2%). **d**, When rejection rate is broken down by offer size, we see that the ASD group ( $n = 56$  participants) rejected a lower percentage of high offers than the two online groups ( $n = 56$  participants each) during the controllable condition (two-sided mixed-effects model with random intercept for matched pair ID,  $P$  values false discovery rate (FDR)-corrected for multiple

comparisons:  $F(122) = 6.12, P = 0.009, \eta_{\text{partial}}^2 = 0.09$ ; mean ASD: 29.1%, mean HT: 45.8%, mean LT: 49.5% (left)). The groups ( $n = 56$  participants each) did not differ in rejection rates in the uncontrollable condition (all two-sided mixed-effects models with random intercepts for matched pair IDs,  $P$  values FDR-corrected for multiple comparisons: low offers:  $F(2,111) = 1.19, P = 0.46, \eta_{\text{partial}}^2 = 0.02$ ; medium offers:  $F(2,111) = 0.46, P = 0.71, \eta_{\text{partial}}^2 = 0.006$ ; high offers:  $F(2,111) = 3.82, P = 0.075, \eta_{\text{partial}}^2 = 0.06$  (right)). **e**, Unlike the online groups ( $n = 56$  participants each), the ASD group ( $n = 56$  participants) did not detect a difference in controllability between the conditions (two-sided mixed-effects models with random intercepts for matched pair IDs:  $F(2,269) = 18.52, P = 2.91 \times 10^{-8}, \eta_{\text{partial}}^2 = 0.12$ ; mean ASD controllable: 45.47, mean HT controllable: 67.45, mean LT controllable: 61.07, mean ASD uncontrollable: 41.79, mean HT uncontrollable: 19.66, mean LT uncontrollable: 24.70). Error bars represent the s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . Panels adapted with permission from: **a**, ref. 27, Elsevier; **b**, ref. 35 under a Creative Commons license CC BY 4.0.

did not differ from each other ( $t(107) = -1.08, P = 0.53$ , estimated difference = -5.2, 95% CI [-16.7, 6.3], Cohen's  $d = -0.20$ ). Such results suggest that, compared with both online groups, the clinically defined ASD sample was less accurate in their ability to detect changes in social controllability. In conjunction with the rejection rate result, these findings suggest that clinically confirmed adults with ASD, but not those defined solely by high autistic traits, showed altered ability to exert influence and perception of their controllability during social interactions.

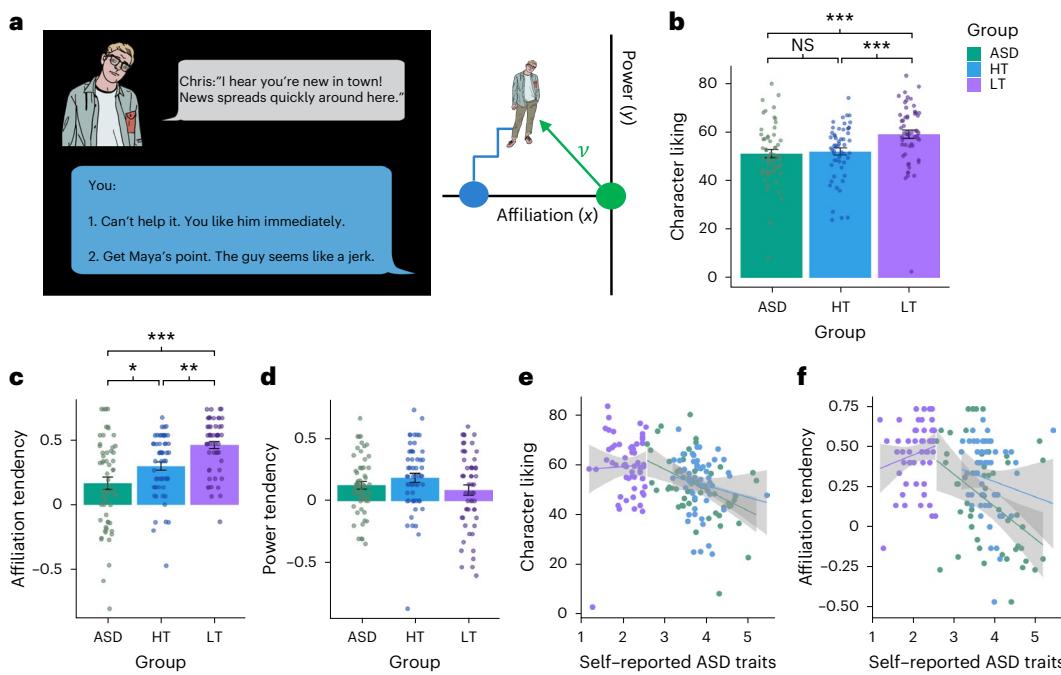
**Social navigation task.** Successful navigation of social interactions requires the ability to dynamically update information about relationship dynamics and the ability to act in accordance with such information. For example, if during a conversation with a stranger, you were to learn that they are actually a close friend of one of your relatives, you may choose to be friendlier with them going forward. ASD is associated with social challenges that may affect the ability to accumulate and/or apply evidence about relationship dynamics and hinder adaptive navigation of social interactions.

To evaluate participants' social feelings and actions during dynamic interactions, we utilized the social navigation task<sup>36</sup>. The social navigation task is a narrative-based game in which participants interact

with a variety of virtual characters with the goal of finding a job and a place to live (Fig. 3a). The task consists of both story-building narrative trials and choice-point interaction trials. During interaction trials, participants choose between one of two ways to interact with a given character. Unbeknownst to the participant, these choices reflect opposing changes in either the power or the affiliation dynamic between them and the characters. At the end of the narrative, participants are asked to rate the characters on how much they liked interacting with them (see Methods for further task details).

We began by investigating participants' subjective feelings toward characters in the task and found that the three groups differed in their ratings of character likability ( $F(2,111) = 8.11, P = 0.0005, \eta_{\text{partial}}^2 = 0.13$ ; Fig. 3b). Compared with the low-trait group, both the high-trait group ( $t(110) = -3.31, P = 0.004$ , estimated difference = -7.1, 95% CI [-12.2, -2.0], Cohen's  $d = -0.63$ ) and the ASD group ( $t(113) = -3.65, P = 0.001$ , estimated difference = -7.9, 95% CI [-13.0, -2.7], Cohen's  $d = -0.69$ ) self-reported reduced liking of characters. The high-trait and ASD groups did not differ in their character liking ( $t(111) = -0.36, P = 0.93$ , estimated difference = -0.77, 95% CI [-5.9, 4.3], Cohen's  $d = -0.07$ ), suggesting comparable subjective experiences.

To explore how each group behaved in the task, we investigated group differences in power and affiliation tendencies, averaged across



**Fig. 3 | Social navigation.** **a**, The social navigation task involved participants interacting with different characters with the goal of finding a job and a home. At each interaction, participants could choose between two options that affected either the affiliation or power dynamics of the relationship. Behind the scenes, each decision would move that character's position accordingly in a social space framed by axes of power and affiliation. **b**, Compared with the low-trait group, the high-trait and ASD groups ( $n = 56$  participants each) both reported a reduced liking of the characters in the social navigation task (two-sided mixed-effects models with random intercepts for matched pair IDs:  $F(2,111) = 8.11$ ,  $P = 0.0005$ ,  $\eta_{\text{partial}}^2 = 0.13$ ; mean ASD: 51.09, mean HT: 51.98, mean LT: 59.10). **c**, Despite having comparable feelings toward characters, the ASD group ( $n = 56$  participants) acted less affiliative than the high-trait group ( $n = 56$  participants; two-sided mixed-effects models with random intercepts for matched pair IDs:  $F(2,111) = 17.21$ ,  $P = 3.098 \times 10^{-7}$ ,  $\eta_{\text{partial}}^2 = 0.24$ ; mean ASD: 0.16, mean HT: 0.30, mean LT: 0.46). **d**, The groups ( $n = 56$  participants each) did not differ in their

power tendencies (two-sided mixed-effects models with random intercepts for matched pair IDs:  $F(2,163) = 1.89$ ,  $P = 0.15$ ,  $\eta_{\text{partial}}^2 = 0.02$ ; mean ASD: 0.13, mean HT: 0.19, mean LT: 0.09). **e**, No group-by-trait interaction on character liking was detected (two-sided mixed-effects models with random intercepts for matched pair IDs:  $F(2,155) = 1.76$ ,  $P = 0.18$ ,  $\eta_{\text{partial}}^2 = 0.02$ ). **f**, However, the relationship between affiliative behavior and self-reported traits differed by group ( $n = 56$  participants each; two-sided mixed-effects models with random intercepts for matched pair IDs:  $F(2,160) = 3.42$ ,  $P = 0.035$ ,  $\eta_{\text{partial}}^2 = 0.04$ ). Only the ASD group ( $n = 56$  participants) showed a negative correlation between self-reported traits and affiliation tendency (two-sided Pearson's correlation,  $P$  value FDR-corrected for multiple comparisons:  $r(54) = -0.38$ ,  $P = 0.01$ ). Error bars represent the s.e.m.; gray shading around regression lines represents the 95% confidence interval. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . Panel **a** adapted with permission from ref. 37, Elsevier.

all characters. A significant three-group difference in affiliation tendency ( $F(2,111) = 17.21$ ,  $P = 3.10 \times 10^{-7}$ ,  $\eta_{\text{partial}}^2 = 0.24$ ; Fig. 3c) revealed that the ASD group acted significantly less affiliative with the characters than both the high-trait group ( $t(111) = -2.63$ ,  $P = 0.026$ , estimated difference = -0.13, 95% CI [-0.25, -0.01], Cohen's  $d = -0.50$ ) and the low-trait group ( $t(112) = -5.86$ ,  $P = 1.45 \times 10^{-7}$ , estimated difference = -0.30, 95% CI [-0.42, -0.18], Cohen's  $d = -1.11$ ), indicating unique behavioral tendencies in the clinically defined sample. The high-trait group was also less affiliative than the low-trait group ( $t(110) = -3.25$ ,  $P = 0.004$ , estimated difference = -0.16, 95% CI [-0.28, -0.04], Cohen's  $d = -0.61$ ). The groups did not differ in their power tendencies ( $F(2,163) = 1.89$ ,  $P = 0.15$ ,  $\eta_{\text{partial}}^2 = 0.02$ ; Fig. 3d).

Last, we explored the relationship between social tendencies and self-reported autistic traits or subjective task ratings in each group. There was no group-by-trait interaction on character liking ( $F(2,155) = 1.76$ ,  $P = 0.18$ ,  $\eta_{\text{partial}}^2 = 0.02$ ; Fig. 3e). Finally, there was a significant group-by-trait interaction on affiliation tendency ( $F(2,160) = 3.42$ ,  $P = 0.035$ ,  $\eta_{\text{partial}}^2 = 0.04$ ; Fig. 3f). While the ASD group showed a negative correlation between self-reported traits and affiliation tendency ( $r(54) = -0.38$ ,  $P = 0.01$ ), there was no relationship in the high-trait ( $r(54) = -0.15$ ,  $P = 0.27$ ) or the low-trait ( $r(54) = 0.18$ ,  $P = 0.27$ ) group. Thus, while the relationship between subjective ratings and self-reported autistic traits did not differ by group membership, the relationship between objective actions and self-reported traits was specific to the clinical sample.

## Self-reported diagnoses

All participants completed a questionnaire asking them to self-report whether they have received certain clinical diagnoses. While the in-person and high-trait online groups reported comparable rates of social anxiety and AVPD diagnoses, only two participants in the high-trait online group reported having ASD diagnoses (see Supplementary Table 3 for further details). In the entire online sample, only ten participants reported having ASD diagnoses. We conducted an additional analysis to evaluate how these ten participants ('online ASD') compared with our in-person ASD sample. We found that the social traits and behaviors of the in-person ASD sample more closely resembled those of the online ASD participants compared with individuals with high autistic traits (see Supplementary Table 4 for further details). Although the small sample size should be interpreted with caution, such findings may suggest that self-reported diagnoses provide more clinically relevant information than self-reported traits. As dimensional approaches continue to gain popularity, researchers may wish to re-emphasize the merits of collecting diagnostic information in conjunction with symptom levels.

## Discussion

Here we sought to investigate the phenotypic similarities and differences between online participants with high self-reported autistic traits and those with an ASD diagnosis confirmed in person via clinician evaluation. We identified a lack of agreement between self-rated and clinician-assessed trait measures, highlighting the need for separate

interpretations of each. When investigating each group's social behavior, although high-trait and clinically ascertained participants had similar levels of self-reported autistic traits, we found that only individuals with confirmed ASD showed impairments in recognizing opportunities to exert social control and reduced affiliation in their interactions with virtual characters; by contrast, high-trait individuals identified online showed comparable social behaviors to low-trait individuals. These results provide a caution for future online research: when attempting to identify and draw overarching conclusions about certain diagnostic groups, self-reported symptom surveys alone may not be sufficient.

Despite the lack of identified measurement agreement in this study, we do not believe that these results suggest that self-report questionnaires are invalid for ASD research. On the contrary, they are important tools for understanding the subjective experiences, levels of internal distress or wellbeing and needs of people with ASD. In the context of well-characterized samples, self-reports are crucial to ensure that individuals with lived experience have a role in shaping the narrative surrounding them, as they can challenge baseless assumptions regarding the intentions or reasoning behind the behaviors of people with ASD. Rather than dismiss the importance of self-views, the results provide a caution for the use of self-report alone for defining or extrapolating about a diagnostic group as a whole.

The discrepancy we detected between self-reported BAPQ and clinician-rated ADOS scores in the in-person ASD group is consistent with previous reports using different measures<sup>11–13</sup>. Discrepancies between self- and observer-rated traits are not uncommon among individuals with altered introspection; they have been reported in a variety of conditions characterized by altered insight, including depression<sup>37</sup> and schizophrenia<sup>38</sup>. Evidence suggests that insight differences in such conditions may be more pronounced in certain domains. Among individuals with schizophrenia, for example, those with reduced insight have been shown to over-report their levels of extroversion but accurately report other personality traits, suggesting insight may have an important role in the reporting of social tendencies specifically<sup>39</sup>. Reduced social self-awareness has been widely reported in ASD<sup>40,41</sup> and likely contributes to discrepancies between self- and clinician report. It is possible that, despite presenting with mild social symptoms to the outside observer, autistic individuals with more social awareness may report greater social difficulties due to increased insight and anxiety regarding their social differences from typically developing peers<sup>40,42</sup>.

In the social controllability task, the ASD group rejected a smaller percentage of high offers in the controllable condition compared the online groups. This reduced rejection of 'good' offers hindered their ability to receive better offers down the line, suggesting they did not take advantage of the controllability offered by the condition. In line with this, we also saw that the ASD group did not self-report any differences in the perceived controllability of the conditions. Such results may stem from reductions in ToM-related understanding of others' motivations in the clinical ASD group but not the high-trait group. To distinguish between random and non-random behaviors on the part of the players, participants must realize that the other players are motivated to receive the largest amount of money possible. To achieve this understanding, one might use prior information (that is, past offers) to build expectations about future behaviors (that is, players will give you repeatedly low offers as long as you continue to accept them) that would fit a given intention (that is, players want to maximize gain) and evaluate their accuracy. In ASD, impaired ability to predict offers and understand players' intentions may lead to a lack of distinction between random and non-random (goal-directed) behavior. Indeed, individuals with ASD display a reduced understanding of social intentions—including whether actions are goal-directed<sup>43</sup>—that appears to stem from a reduced tendency to form expectations based on prior social information<sup>44</sup>. It is also possible that the reduced perception of controllability seen in ASD is caused by impaired affordance

perception, which refers to the ability to ascertain which actions are available for you to take in a given environment. Autistic individuals have been shown to inaccurately estimate action capabilities in the perceptual-motor domain<sup>45</sup>, and such impairments are theorized to extend into the social domain<sup>46</sup>. In any case, the high-trait and low-trait online groups showed comparable behavior across all task measures, suggesting that this impaired detection of others' goal-directed behaviors and/or perception of the actions available to oneself is specific to individuals with a confirmed ASD diagnosis.

In the social navigation task, although both the high-trait and ASD groups reported liking the characters less than the low-trait group, only the ASD group was less affiliative with characters during their interactions than other groups. Such results highlight the importance of measuring behavior for achieving a comprehensive understanding of trait presentation. The high-trait and ASD groups were aligned in their subjective beliefs, about both their traits and their opinions of others, but these beliefs did not translate into comparable social behaviors during the task. Considering that pro-affiliative behavior is often considered to be polite, and that individuals with ASD frequently exhibit lower adherence to social conventions<sup>41</sup>, this difference may be reflective of reduced awareness of or incentive to follow social pleasantries norms in ASD. By contrast, those without a confirmed diagnosis may be more inclined or better able to act friendly despite potentially neutral feelings or dislike of characters. In line with this idea, although we did not detect a group difference in the relationship between character liking and self-reported traits, we detected a distinct relationship between self-reported traits and affiliative behavior only within the ASD group—those with a higher level of traits were the least friendly with the characters. Such results provide further evidence that self-reported traits have different implications in individuals with and without a confirmed ASD diagnosis. Altogether, the findings from both tasks suggest that samples defined by online self-report are phenotypically distinct from clinically ascertained samples and that using such online samples to answer questions about social interaction may not be reflective of ASD as a whole.

In our study, the online group with high autistic traits also self-reported heightened levels of social anxiety and AVPD symptoms compared with the in-person ASD group. In addition, although our online high-trait and in-person ASD groups self-reported similar rates of social anxiety and AVPD diagnoses, less than 5% of individuals in the online high-trait group reported to be clinically diagnosed with ASD. Thus, it is possible that the online group represents individuals with both subclinical and clinical levels of socially avoidant and/or anxious traits without co-occurring ASD. Self-reported autistic traits in the general population, absent ASD diagnoses, may be reflective of generalized social avoidance and self-consciousness regarding social skills rather than autism-specific social difficulties. Supporting the existence of this phenotype, large-scale online studies investigating latent psychiatric factors in the general population have identified transdiagnostic dimensions characterized by similar socially avoidant/anxious traits<sup>25,26</sup>. As we have shown, online participants who report elevated internal perceptions of social difficulties (that is, self-reported emotional or cognitive struggles that others may not notice) as indexed by BAPQ scores also show different social behaviors from those with a clinical diagnosis who show elevated external difficulties (that is, socially inappropriate actions or visible struggles, as described by clinician report), suggesting the diagnosis and the dimension are not synonymous. Although social anxiety commonly co-occurs with ASD anxiety, it is still only represented in less than half of ASD cases<sup>47</sup>, and co-occurrence with AVPD is even less common<sup>48</sup>. It may be the case that self-reported traits lack diagnostic specificity, especially at subclinical thresholds, whereas clinicians are better able to assess symptoms rising to clinical relevance and to assign them the most parsimonious diagnoses through comprehensive analysis of both observed external behaviors and reported experiences.

An important implication of the distinction between diagnosis and traits is that we must be cautious not to extrapolate about the needs of one group on the basis of the findings from research conducted in the other. For example, despite doing reasonably well by external metrics of social abilities, individuals with high self-reported autistic, anxious and avoidant traits may need supports or intervention toward boosting self-confidence and reducing anxiety and negative self-talk rather than social skills training. By contrast, individuals who self-report few traits but present to clinicians with observable difficulties in social interaction may benefit from more skills-focused training to aid in quality-of-life outcomes such as independent living, relationships and employment. This distinction is important because, without it, there is a potential risk of harm (or at least reduced access to benefits) to individuals with ASD who require more behavioral support and access to accommodations; if online self-report-based samples are used to represent the whole diagnostic spectrum despite clear differences in behavior, implications for intervention and accommodation may be biased.

This study should be interpreted with the following limitations in mind. First, we relied on a single self-reported autism trait measure—the BAPQ—because of its strong psychometric properties in both the general population and in those with an ASD diagnosis<sup>17,49,50</sup>. However, other surveys such as the Autism-Spectrum Quotient<sup>51</sup> are also commonly used in research to assess autistic traits, although we note that the Autism-Spectrum Quotient also does not always converge with clinical/caregiver impressions<sup>12,14,16</sup>, similar to the BAPQ–ADOS discrepancy identified in the current study. Second, as our study does not specifically measure insight, we cannot determine whether the discrepancy between self- and clinician-rated symptoms is directly related to insight differences in ASD. Future research should utilize insight paradigms to test this theory. Third, the range of IQ scores in our in-person sample is high, suggesting that our sample may not be representative of the spectrum as a whole. While fluent language capabilities were necessary for completion of our study procedures, future studies with less cognitively demanding tasks should extend the generalizability of this work by including adults with ASD who have lower cognitive functioning and/or higher support needs. In addition, we do not have evidence to examine whether the current findings are specific to ASD or generalizable to other psychiatric diagnoses such as schizophrenia or personality disorders where impaired insight can be a trait. Future research is needed to investigate the broader implications of this work.

Finally, since the inception of this study, Prolific has added a screening tool that allows researchers to specifically select participants that self-report having received a formal clinical diagnosis of ASD. Although this information still lacks the verification and context provided by a full clinical evaluation, the ability to specifically recruit individuals who report a clinical diagnosis would allow researchers to access a larger pool of participants who identify as autistic. Given our preliminary findings in the few individuals who self-report an autism diagnosis in our online sample (Supplementary Table 4), future work should investigate whether the use of additional trait measures and/or selection based on self-reported diagnoses in online studies would identify a group that shows behavior more closely aligned with the ASD phenotype. It is possible that other differences will become apparent; for example, individuals with high socially avoidant and anxious traits may be especially inclined toward participating in non-confrontational online studies rather than in-person studies involving face-to-face interactions, highlighting another important consideration when conducting online research. Future online studies recruiting individuals with self-reported diagnosis should also investigate potential platform differences in social profiles.

As online research continues to proliferate, we must consider the limitations of online approaches when determining which scientific questions they are best suited to answer. Questions about

transdiagnostic traits and symptoms, for example, avoid the issues with diagnostic specificity in self-report and may be well suited for testing with online platforms, especially for traits not associated with impaired insight. Online research is a powerful tool that will continue to help answer important questions in human-participant research. However, the results of the current study suggest that online approaches in psychiatry should be used in tandem with—rather than as a replacement for—lab-based research, and that over-generalization of findings should be avoided in research relying on self-reported traits. For questions that require big data, researchers have other tools at their disposal: pooling resources, developing cross-site collaborations or utilizing resources such as Simons Foundation Powering Autism Research<sup>52</sup> will allow for large-scale replications of lab-based studies in ASD that are less reliant on self-report.

## Methods

### Participants

The study was approved by the institutional review board at the Icahn School of Medicine at Mount Sinai, and all participants provided informed consent before participation. All participants were compensated for their participation. Baseline payment was US\$20 per hour for in-person participants and an average of US\$17.25 total for online participants. All participants were additionally paid a bonus based on the reward drawn from a random trial of the social control task. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Online participants were enrolled in the study as part of a larger online project examining social cognition and mental health. Participants were recruited from Prolific ([www.prolific.co](http://www.prolific.co)), an online research participant recruitment site, with the eligibility criteria of (1) aged between 18 and 64, (2) currently living in the USA and (3) >90% approval rating in Prolific. Participants provided consent by clicking 'I Consent' after reading information about the study and were paid for their participation after completion, in accordance with policies on Prolific and at Mount Sinai's School of Medicine. A total of 1,499 individuals attempted the initial study, which included the social controllability task (April 2020). From this, 14 participants were excluded due to duplication of their data files, and an additional 143 participants were excluded for flat behavior during the task (accepting or rejecting all offers). Of the initial push, 1,269 responded to a follow-up study containing relevant questionnaires (June–August 2020); 38 were excluded for questionnaire non-completion, 9 were excluded for exceeding the questionnaire time limit, and 47 were excluded due to missed attention checks or ID errors. This resulted in a total of 1,041 participants with usable questionnaire and social controllability task data. Also out of the initial push, 733 participants responded to a follow-up study to complete the social navigation task (April 2021–January 2022); 157 were excluded for either not having a plausible average decision response time (within  $\pm 2$  s.d.) or having at- or below-chance post-task memory, resulting in 576 participants with complete social navigation task data. In total, 502 online participants completed all aspects of the study.

Over the course of approximately 3 years, 259 individuals were screened in person for inclusion/exclusion by clinical staff at the Seaver Autism Center for Research and Treatment at the Icahn School of Medicine at Mount Sinai (August 2021–June 2023). Participants were recruited through announcements posted on physical flyers around New York City and email listservs with the eligibility criteria of (1) age between 18 and 50, (2) meet criteria for ASD and (3) IQ > 70. To be inclusive, both those with past ASD diagnoses and those referred or self-referring with questions of autism were screened. Of the 259 participants screened, 171 did not meet criteria for ASD; 88 participants met clinical criteria and enrolled in the current study. Participants were screened for ASD by licensed, research-reliable clinicians using

the ADOS-2, developmental and clinical history, self- and informant (for example, parent or roommate) report of symptoms and adaptive behavior, cognitive functioning and clinical judgment regarding whether the individual meets the criteria for ASD in the *Diagnostic and Statistical Manual of Mental Disorders* 5th edition<sup>60</sup>. Of the 88 initially enrolled in the study, 4 were excluded due to a loss to follow-up and/or unavailability to come into the lab. Of the 84 who attempted the tasks, 64 performed the tasks inside of the magnetic resonance imaging scanner to examine neural questions for an additional study, and 20 performed the tasks outside of the scanner on a laptop due to magnetic resonance imaging contraindications. Both groups were included in this study. To be included in the final sample for the social navigation task, participants had to respond on at least 75% of trials and have above chance post-task memory scores. To be included in the final sample for the social controllability task, participants could not have flat behavior (for example, accepted or rejected all offers in either condition). After exclusion, the final sample for the social navigation task was 71 participants, and the final sample for the social controllability task was 67 participants; 56 successfully completed both tasks without exclusion. The final sample of 56 participants included 18 (32.1%) adults who had received an ASD diagnosis before enrolling in our study, while the other 38 (67.9%) enrolled because they identified as autistic and received a first-time diagnosis through this study.

## Measures

To assess levels of autistic traits in the sample, all participants completed the BAPQ. The BAPQ was selected due to its high sensitivity, specificity and test-retest reliability<sup>17,49,50</sup>. While originally designed to assess autistic traits in the non-autistic relatives of individuals with ASD, the BAPQ's strong psychometric properties<sup>17,49</sup> and lack of ceiling effects<sup>53</sup> in individuals with ASD suggest that it performs well at identifying autistic characteristics in clinical populations<sup>50</sup> as well as the general population. As a result, the BAPQ has evolved into a well-used general measure for autistic traits in populations both with and without ASD<sup>53</sup>. All participants completed additional questionnaires to investigate traits of other psychiatric diagnoses, including Liebowitz Social Anxiety Scale<sup>54</sup> (avoidance questions) and the Avoidant Personality Disorder Impairment Scale<sup>55</sup>. The in-person participants additionally completed the ADOS-2 (module 4), a standard clinical assessment measure for ASD. Total algorithm raw scores for the ADOS were used in our analyses<sup>56</sup>.

To assess self-reported clinical diagnoses, all participants were asked whether they had ever been medically diagnosed with or hospitalized with a series of neurodevelopmental/psychiatric disorders. Space was also provided to write in any unlisted diagnoses.

To assess cognitive ability, all in-person participants completed either the *Wechsler Abbreviated Scale of Intelligence* 2nd edition<sup>57</sup> or the *Wechsler Intelligence Scale for Adults* 4th edition<sup>58</sup>. To compare across participants, we used the full-scale composite scores from whichever test was available. All online participants completed the 16-item International Cognitive Ability Resource test<sup>59</sup>.

## Grouping

The full online sample ( $n = 502$ ) was subdivided into those who scored above the cut-off score on the BAPQ (high-trait,  $n = 168$ ) and those who scored in the bottom 25% on the BAPQ (low-trait,  $n = 121$ ). To minimize potential differences between in-person and online samples, we selected age- and sex-/gender-matched participants from within both high- and low-trait online groups to match the in-person ASD group. To do so, we developed a matching function in R (see data and code availability for a link to the function) that takes each individual in the in-person ASD group and attempts to find a nearest-neighbor match by (1) selecting all remaining, unmatched individuals from the high-trait and low-trait groups that match the participant on gender and (2) selecting the individual from within the newly created matched-gender

groups that is closest in age. If there is no gender match (that is, for the non-binary individuals, of whom there were more in the in-person than online groups) or if gender information is missing, the function matches on sex assigned at birth. This resulted in three groups with 56 participants each: high-trait, low-trait and ASD.

In addition to matched samples, we also assessed differences between our in-person ASD group and the full, unmatched high-trait and low-trait groups. The results remain consistent (see Supplementary Table 2 for further details).

## Experimental paradigms

**Social controllability task.** The social controllability task<sup>26,27,34</sup> investigates how individuals exert control over others to maximize rewards. Participants were paired with virtual players from two 30-person teams, denoted by a town name ('Aldertown' and 'Banyan Bay') as well as a common color for the background and team members' shirts. In each trial, the virtual partner proposed a way to split \$20 (for example, \$8 for you, \$12 for them), and the participant had to decide whether to accept or reject the offer. If the participant chose to accept, both parties received the proposed amounts. If they chose to reject the proposal, neither party received any money. Each team represented a different condition: controllable or uncontrollable. Although the participants were told that they 'may or may not have influence over this team's offers' for both teams, they were not explicitly instructed that they had control over only one team, or which team represented which condition. The order of the conditions was randomized across participants. Importantly, a previous study using this task showed clear differences when participants were instructed that they were 'playing with a computer' instead of 'playing with virtual human partners', suggesting the human version of the task successfully probes social-specific behaviors<sup>34</sup>.

In the controllable condition, participants could either increase the value of the next offer by rejecting the current offer or decrease the value of the next offer by accepting the current offer. The amount of the offer change was determined in a probabilistic manner: 1 in 3 chance of changing the offer by \$2, 1 in 3 chance of changing the offer by \$1, and 1 in 3 chance of no change. By contrast, in the uncontrollable condition, offer amounts were randomly sampled from a predetermined distribution (mean = \$5.00, s.d. = \$2.30), and the order of task conditions was randomized for each participant. In both conditions, the initial offer was \$5, and the offers were constrained to be an integer between \$1 and \$9 (inclusive). At the end of the task, participants were asked to rate how much control they perceived they had over each team on a scale of 0–100%. The task was coded in Psychopy (in-person study, version 2021.1.4) and JavaScript (online study, version ES2019).

**Social navigation task.** The social navigation task<sup>36</sup> is a narrative-based game in which participants interact with a variety of virtual characters. To adapt the original task for use in a clinical population and allow added check-ins as needed, the task was divided into four blocks of roughly equal length, following the natural cut points in the narrative (that is, transition into a new scene). At the start of the game, participants were told they had just moved to a new town and needed to find a job and a place to live. They were asked not to overthink their choices and to behave as they would in real life. The task consisted of narrative trials, which contained images of characters and narrative-progressing text, and decision trials, in which the participant had to choose between two ways of interacting with a given character. To select a choice, participants pressed key 1 or 2 on the computer keyboard or in-scanner button box. Although the task appeared to follow a 'choose your own adventure' style of dynamic storytelling, the slides were the same regardless of participants' decisions. The slides that appeared after the decision trials were written to have narrative continuity regardless of the specific decisions that were made. To minimize the potential for internal biases influencing results, the race (light-versus dark-skinned) and gender (masculine versus feminine presenting) of the characters

were counterbalanced (for in-person participants) or randomized (for online participants) across versions. After the task, participants completed a set of questions, including ratings of how much they liked the characters, as well as a set of memory questions to ensure attention during the task.

Unbeknown to the participant, each decision trial in the task probed choices in either the affiliation or power domain. Affiliation decisions included, for example, whether or not to share physical touch, physical space or information (for example, to share their thoughts on a topic). Power decisions included, for example, whether to submit to versus issue a directive/command, or otherwise exert versus give control. Each option would lead to changes in opposing directions, coded as either +1 or -1 depending on whether it was pro- or anti-affiliative for the affiliation trials, or gave power to the character versus took power away from the character for the power trials. The order of the options within a decision trial was counterbalanced across participants. Over the course of the narrative, participants interacted with five different characters holding a variety of social roles, each with six affiliation decisions and six power decisions, for a total of 60 decisions. There was also a neutral character with three neutral decisions that did not change their social location; these trials were not included in these analyses.

Behind the scenes, participants' choices during the decision trials moved the characters' positions within a two-dimensional social space framed by axes of power and affiliation. Each character started at the origin, with neutral affiliation and power (0,0). With each decision, that character's coordinates were updated in the positive or negative direction along the current dimension. If, for example, the participant chose the pro-affiliative option in an affiliation decision trial, that character would move one unit in the positive direction on the affiliation axis. Thus, at any point in the task, the characters' two-dimensional coordinates were the cumulative sums of the participant's affiliation and power decisions in those specific relationships. To get summary measures of participants' social tendencies, we calculated the means of their decisions in the power and affiliation domains separately for each character and then averaged across characters. The task was coded in Psychopy (in-person study, version 2021.1.4) and JavaScript (online study, version ES2019).

## Statistics

For all analyses evaluating the relationships between continuous variables (task or trait), we utilized general linear models. For all analyses evaluating group differences in traits and task performance, or group-by-trait interaction effects on continuous variables, mixed-effects regression models containing a random intercept for each matched pair were conducted using the lme4 package in R. Degrees of freedom for these analyses were calculated using Satterthwaite's method<sup>61</sup> (the default in lme4). Post hoc pairwise comparisons parsing the direction of group effects were conducted using the emmeans package in R. Estimated marginal means were calculated for each group, and pairwise contrasts were performed with Tukey's adjustment for multiple comparisons to control the family-wise error rate and the Kenward-Roger method<sup>62</sup> to approximate degrees of freedom. Significant trait interactions were followed up by two-tailed Pearson correlations to parse the direction of effects in each group, with *P*-values FDR-corrected for multiple comparisons. All trait and task performance variables were continuous; the group variable was categorical, with the in-person ASD group defined as the reference. All statistical tests in the study controlled for age and sex.

To confirm differences in self-reported autistic traits (as designed through group selection), we tested for differences in BAPQ scores across all three groups. As an exploratory follow-up to further characterize the groups, we also tested for differences in traits of other psychiatric disorders characterized by differences in social behavior: avoidant personality disorder and social anxiety. To investigate

agreement between self-rated and clinician-rated autistic traits, we tested for relationships between BAPQ scores and ADOS scores in the in-person sample. As an exploratory follow-up, we also tested for relationships between corresponding subdomains (restricted and repetitive behaviors: BAPQ 'rigid' subscale and ADOS 'restricted and repetitive behaviors' subdomain; social behavior: averaged BAPQ 'aloof' and 'pragmatic language' subscales and ADOS 'social affect' subdomain). For the social controllability task, we investigated group-by-condition interactions on overall rejection rates and perceived control. To further investigate rejection rate patterns, we evaluated group differences in rejection rates for high (\$7–\$9), medium (\$4–\$6) and low (\$1–\$3) offers, with *P*-values FDR-corrected for multiple comparisons within each condition. For the social navigation task, we investigated group differences in power and affiliation behavioral tendencies, as well as self-report ratings of how much they liked interacting with the characters. Finally, to investigate whether the relationship between traits (for example, BAPQ scores) and subjective task ratings/social behavior differed as a function of group, we investigated group-by-trait and group-by-rating interactions on social navigation task variables.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Data for this study are available at [https://github.com/smbneuro5/Online\\_Vs\\_InPerson\\_AutismResearch](https://github.com/smbneuro5/Online_Vs_InPerson_AutismResearch).

## Code availability

Code for this study is available at [https://github.com/smbneuro5/Online\\_Vs\\_InPerson\\_AutismResearch](https://github.com/smbneuro5/Online_Vs_InPerson_AutismResearch).

## References

- Buhrmester, M. D., Talaifar, S. & Gosling, S. D. An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspect. Psychol. Sci.* **13**, 149–154 (2018).
- Palan, S. & Schitter, C. Prolific.ac—a subject pool for online experiments. *J. Behav. Exp. Finance* **17**, 22–27 (2018).
- Douglas, B. D., Ewell, P. J. & Brauer, M. Data quality in online human-subjects research: comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS ONE* **18**, e0279720 (2023).
- Peer, E., Rothschild, D., Gordon, A., Evernden, Z. & Damer, E. Data quality of platforms and panels for online behavioral research. *Behav. Res. Methods* **54**, 1643–1662 (2022).
- Zorowitz, S., Solis, J., Niv, Y. & Bennett, D. Inattentive responding can induce spurious associations between task behaviour and symptom measures. *Nat. Hum. Behav.* **7**, 1667–1681 (2023).
- Mervis, J. E., Vohs, J. L. & Lysaker, P. H. An update on clinical insight, cognitive insight, and introspective accuracy in schizophrenia-spectrum disorders: symptoms, cognition, and treatment. *Expert Rev. Neurother.* **22**, 245–255 (2022).
- Wels, L. et al. Cognitive insight and introspective accuracy in individuals with bipolar disorder: a scoping review. *Neurosci. Appl.* **2**, 101132 (2023).
- Hood, H. K. et al. Poor insight in obsessive-compulsive disorder: examining the role of cognitive and metacognitive variables. *J. Obsessive Compuls. Relat. Disord.* **23**, 100447 (2019).
- Eack, S. M., Greeno, C. G. & Lee, B.-J. Limitations of the patient health questionnaire in identifying anxiety and depression in community mental health: many cases are undetected. *Res. Soc. Work Pract.* **16**, 625–631 (2006).
- Eaton, W. W., Hall, A. L. F., Macdonald, R. & Mckibben, J. Case identification in psychiatric epidemiology: a review. *Int. Rev. Psychiatry* **19**, 497–507 (2007).

11. Lerner, M. D., Calhoun, C. D., Mikami, A. Y. & De Los Reyes, A. Understanding parent-child social informant discrepancy in youth with high functioning autism spectrum disorders. *J. Autism Dev. Disord.* **42**, 2680–2692 (2012).
12. Bishop, S. L. & Seltzer, M. M. Self-reported autism symptoms in adults with autism spectrum disorders. *J. Autism Dev. Disord.* **42**, 2354–2363 (2012).
13. Johnson, S. A., Filliter, J. H. & Murphy, R. R. Discrepancies between self- and parent-perceptions of autistic traits and empathy in high functioning children and adolescents on the autism spectrum. *J. Autism Dev. Disord.* **39**, 1706–1714 (2009).
14. Mazefsky, C. A., Kao, J. & Oswald, D. P. Preliminary evidence suggesting caution in the use of psychiatric self-report measures with adolescents with high-functioning autism spectrum disorders. *Res. Autism Spectr. Disord.* **5**, 164–174 (2011).
15. Sandercock, R. K., Lamarche, E. M., Klinger, M. R. & Klinger, L. G. Assessing the convergence of self-report and informant measures for adults with autism spectrum disorder. *Autism* **24**, 2256–2268 (2020).
16. Ashwood, K. L. et al. Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychol. Med.* **46**, 2595–2604 (2016).
17. Hurley, R., Losh, M., Parlier, M., Reznick, J. & Piven, J. The broad autism phenotype questionnaire. *J. Autism Dev. Disord.* **37**, 1679–1690 (2007).
18. Murray, A. L., Booth, T., McKenzie, K., Kuenssberg, R. & O'Donnell, M. Are autistic traits measured equivalently in individuals with and without an autism spectrum disorder? An invariance analysis of the autism spectrum quotient short form. *J. Autism Dev. Disord.* **44**, 55–64 (2014).
19. Tonge, N. A., Rodebaugh, T. L., Fernandez, K. C. & Lim, M. H. Self-reported social skills impairment explains elevated autistic traits in individuals with generalized social anxiety disorder. *J Anxiety Disord.* **38**, 31–36 (2016).
20. Nobili, A. et al. Autistic traits in treatment-seeking transgender adults. *J. Autism Dev. Disord.* **48**, 3984–3994 (2018).
21. De Cagna, F., Squillari, E., Rocchetti, M. & Fusar-Poli, L. in *Psychopathology in Adolescents and Adults with Autism Spectrum Disorders* (ed. Keller, R.) 157–174 (Springer, 2019).
22. Simon, A. E., Keller, P. & Cattapan, K. Commentary about social avoidance and its significance in adolescents and young adults. *Psychiatry Res.* **297**, 113718 (2021).
23. Hummelen, B., Wilberg, T., Pedersen, G. & Karterud, S. The relationship between avoidant personality disorder and social phobia. *Compr. Psychiatry* **48**, 348–356 (2007).
24. White, S. W., Bray, B. C. & Ollendick, T. H. Examining shared and unique aspects of social anxiety disorder and autism spectrum disorder using factor analysis. *J. Autism Dev. Disord.* **42**, 874–884 (2012).
25. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* **5**, e11305 (2016).
26. Banker, S. M. et al. Disrupted computations of social control in individuals with obsessive-compulsive and misophonia symptoms. *iScience* **25**, 104617 (2022).
27. Na, S. et al. Computational mechanisms underlying illusion of control in delusional individuals. *Schizophr. Res.* <https://doi.org/10.1016/j.schres.2022.01.054> (2022).
28. Lord, C. et al. *Autism Diagnostic Observation Schedule Second Edition (ADOS-2) Manual (Part 1): Modules 1–4* (Western Psychological Services, 2012).
29. Kimhi, Y. Theory of mind abilities and deficits in autism spectrum disorders. *Top. Lang. Disord.* **34**, 329 (2014).
30. Tager-Flusberg, H. Evaluating the theory-of-mind hypothesis of autism. *Curr. Dir. Psychol. Sci.* **16**, 311–315 (2007).
31. Woodcock, K. A., Cheung, C., González Marx, D. & Mandy, W. Social decision making in autistic adolescents: the role of theory of mind, executive functioning and emotion regulation. *J. Autism Dev. Disord.* **50**, 2501–2512 (2020).
32. Keifer, C. M., Mikami, A. Y., Morris, J. P., Libsack, E. J. & Lerner, M. D. Prediction of social behavior in autism spectrum disorders: explicit versus implicit social cognition. *Autism* **24**, 1758–1772 (2020).
33. Cannon, J., O'Brien, A. M., Bungert, L. & Sinha, P. Prediction in autism spectrum disorder: a systematic review of empirical evidence. *Autism Res.* **14**, 604–630 (2021).
34. Na, S. et al. Humans use forward thinking to exploit social controllability. *eLife* **10**, e64983 (2021).
35. Na, S., Rhoads, S. A., Yu, A. N. C., Fiore, V. G. & Gu, X. Towards a neurocomputational account of social controllability: from models to mental health. *Neurosci. Biobehav. Rev.* **148**, 105139 (2023).
36. Tavares, R. M. et al. A map for social navigation in the human brain. *Neuron* **87**, 231–243 (2015).
37. Enns, M. W., Larsen, D. K. & Cox, B. J. Discrepancies between self and observer ratings of depression: the relationship to demographic, clinical and personality variables. *J. Affect. Disord.* **60**, 33–41 (2000).
38. Engel, M. & Lincoln, T. M. Concordance of self- and observer-rated motivation and pleasure in patients with negative symptoms and healthy controls. *Psychiatry Res.* **247**, 1–5 (2017).
39. Bell, M., Fiszdon, J., Richardson, R., Lysaker, P. & Bryson, G. Are self-reports valid for schizophrenia patients with poor insight? Relationship of unawareness of illness to psychological self-report instruments. *Psychiatry Res.* **151**, 37–46 (2007).
40. Huang, A. X. et al. Understanding the self in individuals with autism spectrum disorders (ASD): a review of literature. *Front. Psychol.* **8**, 1422 (2017).
41. Lombardo, M. V. & Baron-Cohen, S. The role of the self in mindblindness in autism. *Conscious. Cogn.* **20**, 130–140 (2011).
42. Hotton, M. & Coles, S. The effectiveness of social skills training groups for individuals with autism spectrum disorder. *Rev. J. Autism Dev. Disord.* **3**, 68–81 (2016).
43. Ganglmaier, K., Schuwerk, T., Sodian, B. & Paulus, M. Do children and adults with autism spectrum condition anticipate others' actions as goal-directed? A predictive coding perspective. *J. Autism Dev. Disord.* **50**, 2077–2089 (2020).
44. Chambon, V. et al. Reduced sensitivity to social priors during action prediction in adults with autism spectrum disorders. *Cognition* **160**, 17–26 (2017).
45. Linkenauger, S. A., Lerner, M. D., Ramenzoni, V. C. & Proffitt, D. R. A perceptual–motor deficit predicts social and communicative impairments in individuals with autism spectrum disorders. *Autism Res.* **5**, 352–362 (2012).
46. Hellendoorn, A. Understanding social engagement in autism: being different in perceiving and sharing affordances. *Front. Psychol.* **5**, 850 (2014).
47. Spain, D., Zivrali Yarar, E. & Happé, F. Social anxiety in adults with autism: a qualitative study. *Int. J. Qual. Stud. Health Well-Being* **15**, 1803669 (2020).
48. Rinaldi, C., Attanasio, M., Valenti, M., Mazza, M. & Keller, R. Autism spectrum disorder and personality disorders: comorbidity and differential diagnosis. *World J. Psychiatry* **11**, 1366–1386 (2021).
49. Sasson, N. J. et al. The broad autism phenotype questionnaire: prevalence and diagnostic classification. *Autism Res.* **6**, 134–143 (2013).
50. Nishiyama, T. et al. Comprehensive comparison of self-administered questionnaires for measuring quantitative autistic traits in adults. *J. Autism Dev. Disord.* **44**, 993–1007 (2014).

51. 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. *J. Autism Dev. Disord.* **31**, 5–17 (2001).
52. SPRAK Consortium. SPARK: a US cohort of 50,000 families to accelerate autism research. *Neuron* **97**, 488–493 (2018).
53. Hull, L. et al. Development and validation of the camouflaging autistic traits questionnaire (CAT-Q). *J. Autism Dev. Disord.* **49**, 819–833 (2019).
54. Liebowitz, M. R. Social phobia. *Mod. Probl. Pharmacopsychiatry* **22**, 141–173 (1987).
55. Liggett, J., Carmichael, K. L. C., Smith, A. & Sellbom, M. Validation of self-report impairment measures for Section III obsessive-compulsive and avoidant personality disorders. *J. Pers. Assess.* **99**, 1–14 (2017).
56. Hus, V. & Lord, C. The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. *J. Autism Dev. Disord.* **44**, 1996–2012 (2014).
57. Wechsler, D. *Wechsler Abbreviated Scale of Intelligence* 2nd edn (Pearson, 2011).
58. Wechsler, D. *Wechsler Adult Intelligence Scale* 4th edn (Pearson, 2008).
59. Condon, D. M. & Revelle, W. The international cognitive ability resource: development and initial validation of a public-domain measure. *Intelligence* **43**, 52–64 (2014).
60. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (American Psychiatric Association, 2013).
61. Gaylor, D. W. & Hopper, F. N. Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics* **11**, 691–706 (1969).
62. Kenward, M. G. & Roger, J. H. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983–997 (1997).

## Acknowledgments

S.B. is supported by the Seaver Autism Center for Research and Treatment (graduate fellowship) and R01MH122611. J.H.F.-F. is supported by R01MH122611, R01MH119172, R01NS105845 and SFARI award 877761. X.G. is supported by the National Institute of Mental Health (grants R21MH120789, R01MH122611, R01MH123069), and SFARI award 877761. D.S. is supported by the National Institute of Mental Health (grants R21MH120789, R01MH122611, R01MH123069) and SFARI award 877761. Online data collection was supported by internal funding from the ISMMS. This work was supported in part through the computational and data resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences.

## Author contributions

S.M.B. and X.G. conceptualized the study. S.M.B., D.S., J.H.F.-F. and X.G. determined the proper methodology for the study. M.S., S.N., M. Heflin, S.B., J.T., A.W.P. and A.A.T. collected the data. S.M.B. and M. Harrington performed data analysis. S.M.B. and M. Harrington wrote the original draft of the manuscript. S.M.B., M. Harrington, M.S., D.S., J.H.F.-F. and X.G. reviewed and edited the final manuscript. D.S., J.H.F.-F. and X.G. acquired funding for the research. S.N and M.S. provided task and coding resources. D.S., J.H.F.-F. and X.G. supervised the project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44220-025-00385-8>.

**Correspondence and requests for materials** should be addressed to Sarah M. Banker or Xiaosi Gu.

**Peer review information** *Nature Mental Health* thanks Melanie Penner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection The tasks were coded and collected using Psychopy (in-person study, version 2021.1.4) and JavaScript (online study, version ES2019).

Data analysis Data analysis was performed in R, version 4.4.0. Statistical packages used included lme4 (version 1.1-35.5) and emmeans (version 1.10.3).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data for this study is available at [https://github.com/smbneuro5/Online\\_Vs\\_InPerson\\_AutismResearch](https://github.com/smbneuro5/Online_Vs_InPerson_AutismResearch)

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

### Reporting on sex and gender

Sex and gender information was collected as a part of this study. Sex and gender were used to match online research participants to in-person research participants, but were not otherwise variables of interest in this study. The sex and gender breakdown of each group is presented in Table 1 of the manuscript.

### Population characteristics

In-person participants (n=56) all met clinical criteria for Autism Spectrum Disorder (ASD), and included 18 (32.1%) adults who had received an ASD diagnosis prior to enrolling in our study, and 38 (67.9%) adults who enrolled because they identified as autistic and received a first-time diagnosis through this study. Online participants (n=502) were adults recruited from the general population with no diagnostic requirements. The online sample was subdivided into those who scored above the cutoff score on the Broad Autism Phenotype Questionnaire (BAPQ; "high-trait", n=168) and those who scored the bottom 25% on BAPQ scores ("low-trait", n=121). From within each of these groups, 56 age- and sex/gender- matched participants were selected to match the in-person ASD sample. This resulted in three groups with 56 participants each: high-trait, low-trait, and ASD. The characteristics for the final groups were as follows: ASD: age mean = 28.07, age standard deviation = 8.53, 51.8% assigned female at birth (n=29); High-trait: age mean = 29.14, age standard deviation = 7.49, 46.4% assigned female at birth (n=26); Low-trait: age mean = 29.88, age standard deviation = 7.90, 44.6% assigned female at birth (n=25). See "behavioural & social sciences study design" for further information.

### Recruitment

Online participants were enrolled in the study as part of a larger online project examining social cognition and mental health. Participants were recruited from Prolific ([www.prolific.co](http://www.prolific.co)), an online research participant recruitment site, with the eligibility criteria of (1) age between 18 and 64, (2) currently living in the United States, (3) >90% approval rating in Prolific. In-person participants were recruited through announcements posted on physical flyers around New York City and email listservs with the eligibility criteria of (1) age between 18 and 50, (2) meet criteria for ASD, (3) IQ >70. In-person participants were screened in-person for inclusion/exclusion by clinical staff at the Seaver Autism Center for Research and Treatment at the Icahn School of Medicine at Mount Sinai. To be inclusive, both those with past ASD diagnoses and those referred or self-referring with questions of autism were screened. As our study offered in-person participants a free evaluation for ASD, many of our in-person participants (n=38, 67.9%) came to the study through self-identification, with no prior clinical diagnosis. As girls/women on the autism spectrum are often under-recognized due to masking tendencies or awareness bias, it's possible that the increased number of female participants in our sample compared to prior studies (and compared to reported national prevalence) is due to a greater tendency for women to seek a diagnosis that may have been overlooked when they were children. Of the 38 of our participants that received a first-time diagnosis through our study, 26 of them (68.4%) were assigned female at birth. However, due to careful matching of the groups on sex and gender, we do not believe that this potential bias should impact the group comparisons outlined in the study.

### Ethics oversight

The study was approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai. Online study: determined exempt; IRB-18-01301. In-person study: IRB-16-01089.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](http://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Study description

This was a quantitative, cross-sectional study examining differences in social behavior and phenotypic profiles between adults recruited online who self-reported a high level of autism symptoms and adults recruited in-person with a confirmed clinical diagnosis of Autism Spectrum Disorder.

### Research sample

While allowing for rapid recruitment of large samples, online research relies heavily on participants' self-reports of neuropsychiatric traits, foregoing the clinical characterizations available in laboratory settings. Autism spectrum disorder (ASD) research is one example for which the clinical validity of such an approach remains elusive. We thus chose to recruit and compare in-person participants with ASD, evaluated by clinicians, and online participants, evaluated through self-reported autistic traits. The online sample consisted of adults recruited from Prolific ([www.prolific.co](http://www.prolific.co)), an online research participant recruitment site, with the eligibility criteria of (1) age between 18 and 64, (2) currently living in the United States, (3) >90% approval rating in Prolific. These participants were unselected for any diagnosis. We followed common procedures for online research recruitment to obtain a sample representative of online studies. In-person participants were screened in-person for inclusion/exclusion by clinical staff at the Seaver Autism Center for Research and Treatment at the Icahn School of Medicine at Mount Sinai. To be inclusive, both those with past ASD diagnoses and those referred or self-referring with questions of autism were screened. The range of IQ scores in our in-person sample is quite high, suggesting that our sample may not be representative of the spectrum as a whole. A full breakdown of the

demographics of each group is available in Table 1 of the manuscript. Participants were not specifically sampled to reflect nationally representative demographics.	
Sampling strategy	Volunteer sampling was used. Sample sizes were determined by power analyses of on target effect sizes based on prior studies using the same paradigms, as well as by what is typically employed in the field.
Data collection	Online: All data was collected remotely on a computer with no experimenter present. In-person: All survey and task data was collected on a computer with just the participant present (researchers left the room after explanation and practice). Autism Diagnostic Observation Schedule (ADOS) was collected through standard procedures and scores were recorded by clinicians on paper forms; scores were then double entered into the RedCap database. Researchers were not blinded.
Timing	Online data was collected through a larger online project examining social cognition and mental health from April-August 2020. In-person data was collected from August 2021 - June 2023.
Data exclusions	Online data: A total of 1,499 individuals attempted the initial study, which included the social controllability task (April 2020). From this, 14 participants were excluded due to duplication of their data files and an additional 143 participants were excluded for flat behavior during the task (accepting or rejecting all offers). Of the initial push, 1,269 responded to a follow-up study containing relevant questionnaires (June - August 2020); 38 were excluded for questionnaire non-completion, 9 were excluded for exceeding the questionnaire time limit, and 47 were excluded due to missed attention checks or ID errors. This resulted in a total of 1,041 participants with usable questionnaire and social controllability task data. Also out of the initial push, 733 participants responded to a follow-up study to complete the social navigation task (April 2021 - January 2022); 157 were excluded for either not having a plausible average decision response time (within +/- 2 standard deviations of mean) or having at- or below-chance post-task memory, resulting in 576 participants with complete social navigation task data. In total, 501 online participants completed all aspects of the study.  In-person data: Of the 259 participants screened, 171 did not meet criteria for ASD; 88 participants met clinical criteria and enrolled in the current study. Participants were screened for ASD by licensed, research reliable clinicians using the ADOS-2, developmental and clinical history, self and informant (e.g., parent, roommate) report of symptoms and adaptive behavior, cognitive functioning, and clinical judgment regarding whether the individual meets DSM-5 criteria for ASD. Of the 88 initially enrolled in the study, 4 were excluded due to a loss to follow-up and/or unavailability to come into the lab. Of the 84 who attempted the tasks, 64 performed the tasks inside of the MRI to examine neural questions for an additional study, and 20 performed the tasks outside of the scanner on a laptop due to MRI contraindications. Both groups were included in this study. To be included in the final sample for the social navigation task, participants had to respond on at least 75% of trials and have above chance post-task memory scores. To be included in the final sample for the social controllability task, participants could not have flat behavior (e.g., rejected or rejected all offers in either condition). After exclusion, the final sample for the social navigation task was 71 participants and the final sample for the social controllability task was 67 participants; 56 successfully completed both tasks without exclusion. The final sample of 56 participants included 18 (32.1%) adults who had received an ASD diagnosis prior to enrolling in our study, while the other 38 (67.9%) enrolled because they identified as autistic and received a first-time diagnosis through this study.
Non-participation	No online participants dropped out or declined participation. Some were "loss to follow up" between the initial and follow-up pushes, as outlined above. Two of the in-person participants were loss-to-follow-up, and two dropped out because they were unable to come into the lab in-person to complete all study procedures.
Randomization	Participants were grouped based on the recruitment platform (online vs in-person). The online participants were subdivided into those who scored above the cutoff score on the BAPQ ("high-trait", n=168) and those who scored the bottom 25% on BAPQ scores ("low-trait", n=121). To minimize potential differences between in-person and online samples, we selected age- and sex/gender-matched participants from within both high- and low-trait online groups to match the in-person ASD group. To do so, we developed a matching function in R (see data and code availability for a link to the function) that takes each individual in the in-person ASD group and attempts to find a nearest-neighbor match by 1) selecting all remaining, unmatched individuals from the high-trait and low-trait groups that match the participant on gender, and 2) selecting the individual from within the newly created matched-gender groups that is closest in age. If there is no gender match (i.e., for the non-binary individuals, of whom there were more in the in-person than online groups) or if gender information is missing, the function matches on sex assigned at birth. This resulted in three groups with 56 participants each: high-trait, low-trait, and ASD.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Involved in the study                                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |

### Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involved in the study                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |