ARTICLE

A neuronal social trait space for first impressions in the human amygdala and hippocampus

Runnan Cao[1]✉, Chujun Lin[2], Johnie Hodge[3], Xin Li[1], Alexander Todorov[4], Nicholas J. Brandmeir[3,5] and Shuo Wang[1,5,6]✉

People instantaneously evaluate faces with significant agreement on evaluations of social traits. However, the neural basis for such rapid spontaneous face evaluation remains largely unknown. Here, we recorded from 490 neurons in the human amygdala and hippocampus and found that the neuronal activity was associated with the geometry of a social trait space. We further investigated the temporal evolution and modulation on the social trait representation, and we employed encoding and decoding models to reveal the critical social traits for the trait space. We also recorded from another 938 neurons and replicated our findings using different social traits. Together, our results suggest that there exists a neuronal population code for a comprehensive social trait space in the human amygdala and hippocampus that underlies spontaneous first impressions. Changes in such neuronal social trait space may have implications for the abnormal processing of social information observed in some neurological and psychiatric disorders.

Molecular Psychiatry; https://doi.org/10.1038/s41380-022-01583-x

INTRODUCTION

Faces are among the most important visual stimuli we perceive and they often convey a wealth of information. When we see a person's face, we can easily recognize their unique identity and general features such as sex and age. The gestalt of facial processing enables us to automatically evaluate faces on multiple trait dimensions (e.g., trustworthiness) [1], and these evaluations predict important social outcomes, ranging from electoral success to sentencing decisions [2]. However, a central challenge in face research is to understand how the brain evaluates faces in general and forms rapid spontaneous impressions of faces on multiple trait dimensions.

It has been shown that neurons in the primate inferotemporal (IT) cortex encode a face space of low-level features, demonstrating a comprehensive neural code for physical variations in faces such as eye shape and skin tone [3–5]. On the other hand, the human amygdala and hippocampus play critical roles in social perception [6, 7] and encode various social trait judgments of faces (i.e., judgments of an individual's temporally stable characteristics). For example, a lesion study agrees with the hypothesis that the amygdala is necessary for judging facial trustworthiness [8], which is further supported by functional neuroimaging studies [9]. We previously utilized single-neuron recordings in the human amygdala to show that the amygdala parametrically encodes facial emotions [10], which are known to shape various social trait judgments of faces such as personality traits [11]. Prior studies have characterized the neural bases of only a few individual trait judgments; however, humans use hundreds of different trait words to describe spontaneous trait judgments of faces [12–14] and automatically evaluate faces on multiple trait dimensions simultaneously. Whether the amygdala and hippocampus encode a comprehensive space for social trait judgments of faces has not yet been determined.

People with autism spectrum disorder (ASD) demonstrate pervasive impairments in face processing and social evaluations of faces [15–19]. An overarching hypothesis is that people with ASD have atypical neural representations of faces, with consequences for abnormal social evaluation, emotion processing, eye movements, recognition, and memory of faces [20]. This hypothesis is supported by differences in brain activity when people with ASD view faces, and a specific neural structure underlying face processing deficits in ASD is the amygdala [21, 22]. For example, single-neuron recordings in the human amygdala show weaker responses to eyes in people with ASD [23], and neuroimaging studies show that amygdala-mediated orientation toward the eyes seen in BOLD-fMRI is dysfunctional in ASD [24]. Yet, it remains to be tested whether abnormal processing of social information from faces in ASD can be explained by an abnormal comprehensive neural representation of social traits.

In this study, we hypothesize there exists a neuronal social trait space in the human amygdala and hippocampus that underlies spontaneous first impressions of faces. Primate research on face processing supports such a possibility: neurons from the macaque temporal lobe encode a multi-dimensional face feature space [3–5], as well as a multitude of social information (for a review see [25]), providing plausible neural mechanisms supporting different dimensions of complex social evaluations. Furthermore, our recent neuroimaging data suggests that the human amygdala encodes

[1]Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA. [2]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA. [3]Department of Neurosurgery, West Virginia University, Morgantown, WV 26506, USA. [4]Booth School of Business, University of Chicago, Chicago, IL 60637, USA. [5]Rockefeller Neuroscience Institute, West Virginia University, Morgantown, WV 26506, USA. [6]Department of Radiology, Washington University in St. Louis, St. Louis, MO 63110, USA. ✉email: runnan.cao@mail.wvu.edu; shuowang@wustl.edu

Received: 28 August 2021 Accepted: 12 April 2022
Published online: 08 June 2022

physical variations in faces (e.g., shape and skin tone) that underlie representations of various social traits [26]. A recent psychological study using the largest number of representatively sampled social traits to date has characterized a comprehensive space for social trait judgments of faces—a four-dimensional space with dimensions interpreted as warmth, competence, femininity, and youth [14]. Based on this comprehensive social trait space, the present study investigated whether there exists a population code (i.e., neuronal population activity collectively contributes to the judgments) for evaluating multimodal social traits in the human amygdala and hippocampus, which will provide the neural basis for first impressions of faces. We also provide a direct replication of our results using an additional dataset and another well-established social trait space. We lastly investigate the behavioral consequence of the neuronal social trait space for those with ASD.

MATERIALS AND METHODS

Patients

A total of 16 single-neuron recording sessions were conducted with 5 patients (4 female) who had undergone surgery to have electrodes implanted to treat intractable epilepsy (see Supplementary Table 1 for patient demographics, epilepsy diagnosis and treatment, and the number of neurons recorded from each brain area and each session; see Supplementary Fig. 1 for electrode locations). The sample size was determined based on our previous experience; and no patients were excluded from the study. All patients provided written informed consent using procedures approved by the Institutional Review Board of West Virginia University (IRB number: 1709745061).

Experimental procedure

We used a 1-back task for the CelebA stimuli (see Supplementary Methods for details). In each trial, a single face was presented at the center of the screen for a fixed duration of 1 s, with uniformly jittered inter-stimulus-interval of 0.5–0.75 s (see Supplementary Methods for details). For the FaceGen stimuli (see Supplementary Methods for details), patients performed two face judgment tasks. In each task, there was a judgment instruction, i.e., patients judged how trustworthy or how dominant a face was. We used a 1–4 scale: "1": not trustworthy/dominant at all, "2": somewhat trustworthy/dominant, "3": trustworthy/dominant, and "4": very trustworthy/dominant. Each image was presented for 1.5 s at the center of the screen.

Online rating of social traits

For the CelebA stimuli, we acquired social trait ratings of the faces from both patients and a large number of participants from the general population (age ($M = 26.20$ years, SD = 7.11), 180/501 females; see Supplementary Table 2 for demographics). Participants were asked to rate the faces on eight social traits using a seven-point Likert scale through an online rating task. The social traits included warm, critical, competent, practical, feminine, strong, youthful, and charismatic, representing the four core psychological dimensions of comprehensive trait judgments of faces (warmth, competence, femininity, and youth; 2 traits per dimension); and these social traits were well validated in a previous study [14] (see Supplementary Methods for more details).

Patients completed the social trait rating task online after they were discharged from the hospital following surgery to treat intractable epilepsy. Three patients completed the rating task and provided ratings for 2 to 5 photos per identity per social trait. Participants from the general population completed the rating task using the Prolific online research platform (see Supplementary Methods for more details and exclusion criteria). We repeated the same procedure to acquire ratings from participants with ASD. In addition, both participants with ASD and controls were asked to provide demographic information and complete the online Autism Spectrum Quotient (AQ) and Social Responsiveness Scale-2 Adult Self Report (SRS-A-SR) questionnaires.

Single-neuron response

Detailed electrophysiology is described in Supplementary Methods (see also Supplementary Fig. 2). Only units with an average firing rate of at least 0.15 Hz during the entire task were considered. Only single units were considered. Trials were aligned to stimulus onset. We used the mean firing rate in a time window 250 to 1250 ms after stimulus onset as the response to each face. Firing rate was then normalized by dividing the mean activity in the baseline (−250 to 0 ms relative to stimulus onset). Such normalization was applied in previous studies that analyzed the similarity between single-neuron responses to visual categories [27].

Face-responsive neurons were identified by comparing the response to faces (i.e., the mean firing rate in a time window 250 to 1250 ms after stimulus onset) to baseline (i.e., −250 to 0 ms relative to stimulus onset) using a two-tailed paired t-test with $p < 0.05$.

Data analysis

For representational similarity analysis (RSA) [28], dissimilarity matrices (DMs) are symmetrical matrices of dissimilarity between all pairs of face images or face identities. In a DM, larger values represent larger dissimilarity of pairs, such that the smallest value possible is the similarity of a condition to itself (dissimilarity of 0). We used the Pearson correlation to calculate DMs (ratings were z-scored and firing rates were normalized to the mean baseline of each neuron), and we used the Spearman correlation to calculate the correspondence between the DMs (Spearman correlation was used because it does not assume a linear relationship [29]; Fisher z-transformation was performed on Pearson's r to ensure that sample distribution was approximately normal). We further used permutation tests with 1000 runs to assess the significance of the correspondence between the social trait DM and the neural response DM. Because the consistency between face images for the same face identity in both social trait ratings and neural responses could inflate the correspondence between the social trait DM and the neural response DM, we averaged the social trait ratings or neural responses across face images for each face identity and calculated the DM between face identities. We further used a moving window (bin size = 500 ms, step size = 50 ms) to measure temporal dynamics. The first bin started −500 ms relative to trial onset (bin center was thus 250 ms after trial onset), and we tested 31 consecutive bins (the last bin was thus from 1000 to 1500 ms after trial onset).

We used a bootstrap with 1000 runs to estimate the distribution of DM correspondence for each participant group. In each run, 70% of the data were randomly selected from each participant group and we calculated the correspondence (Spearman's ρ) between the social trait DM and the neural response DM for each participant group. We then created a distribution of DM correspondence for each participant group, and we compared the mean of the ASD distribution to the control distribution and vice versa to derive statistical significance.

We further used a permutation test with 1000 runs to statistically compare the DM correspondence between participants with ASD and controls. In each run, we shuffled the participant labels and calculated the difference in DM correspondence between participant groups. We then compared the observed difference in DM correspondence between participant groups with the permuted null distribution to derive statistical significance.

Encoding and decoding models are described in Supplementary Methods.

RESULTS

Constructing a comprehensive social trait space

Neurosurgical patients undergoing single-neuron recordings (Supplementary Table 1; see Supplementary Fig. 1 for electrode locations) viewed 500 natural face images of 50 celebrities (10 images per celebrity) while performing a simple one-back task (accuracy = 75.7 ± 5.28% (mean ± SD across sessions)). Additionally, we acquired consensus social trait ratings for the same face stimuli on eight traits from a large population of participants recruited via an online platform (see Methods; 415.75 ± 11.42 (mean ± SD) raters per trait; Supplementary Table 2 and Supplementary Fig. 3a). The eight traits (warm, critical, competent, practical, feminine, strong, youthful, and charismatic) were selected to represent the four comprehensive psychological dimensions of social trait judgments of faces (warmth, competence, femininity, and youth; two traits per dimension; see "Methods") [14]. The inter-rater consistency of these ratings (Supplementary Fig. 3b, c) was comparable to the established study [14] (see also Supplementary Fig. 3d for correlations

between ratings from different modules [each module contained one face image per identity] and Supplementary Fig. 3e for correlations between social traits).

We used the average ratings across participants per face on the eight traits to construct a "social trait space". We verified that this social trait space reproduced the four comprehensive dimensions of facial social trait judgments found in the prior study [14] (Supplementary Table 3). We found that this social trait space demonstrated an organized structure after projecting it onto a two-dimensional space for visualization using t-distributed stochastic neighbor embedding (t-SNE): different images of the same person were clustered, and the two t-SNE dimensions showed the change in two of the four comprehensive psychological dimensions (warmth and femininity) as expected. The trait judgment was highly consistent for different images of the same person. It is worth noting that patients' ratings were consistent with the consensus ratings (Supplementary Results and Supplementary Fig. 4) so we used the consensus ratings for further analysis (see also Supplementary Discussion).

The neuronal population in the amygdala and hippocampus encode the social trait space

We recorded from 490 neurons in the amygdala and hippocampus of 5 neurosurgical patients (16 sessions in total; overall firing rate greater than 0.15 Hz; see Supplementary Fig. 1 for recording locations), which included 242 neurons from the amygdala, 186 neurons form the anterior hippocampus, and 62 neurons from the posterior hippocampus (see Supplementary Table 1 for a breakdown of each individual session; see Supplementary Fig. 2 for assessment of spike sorting quality). We aligned neuronal responses at stimulus onset and used the mean normalized firing rate in a time window from 250 to 1250 ms after stimulus onset for subsequent analyses.

To investigate whether the neuronal population encoded the comprehensive social trait space, we calculated DMs between face identities for social traits (Fig. 1a left; using ratings for eight social traits) and neural responses (Fig. 1a right; using the mean normalized firing rate of neurons), and we assessed the correspondence between the social trait DM and the neural response DM using RSA [28]. We found that the DM from face-responsive neurons (i.e., neurons that had a significant change in firing rate after stimulus onset compared to baseline; see "Methods"; $n = 74$) was significantly correlated with the social trait DM (Fig. 1b; permutation $p < 0.001$), and this was the case for both amygdala neurons (Fig. 1c; $n = 36$, permutation $p = 0.011$) and hippocampal neurons (Fig. 1d; $n = 38$, permutation $p = 0.004$).

We further investigated the impact of race on social trait perceptions. We found that the amygdala and hippocampal neurons encoded the social trait space constructed with Caucasian faces only (Fig. 1e; permutation $p < 0.001$) or Black faces only (Fig. 1f; permutation $p = 0.003$), suggesting that encoding of social traits in the amygdala and hippocampus was independent of racial differences (note that we used Bonferroni correction to control for multiple comparisons in Fig. 1b–g). In addition, we investigated the time course of the correspondence between the social trait DM and the neural response DM (Fig. 1h; corrected for multiple comparisons using false discovery rate [30]). We found that encoding of the social trait space peaked at between 200 to 400 ms after stimulus onset. The response from hippocampal neurons peaked earlier (at ~150 ms; in comparison to ~400 ms for amygdala neurons) and was greater than that from amygdala neurons, but amygdala neurons had a more sustained response than hippocampal neurons (Fig. 1h). We further conducted several control analysis to confirm our results (Supplementary Results; Supplementary Figs. 5 and 6).

Together, our results suggest that the neuronal population in the amygdala and hippocampus collectively encode the geometry of the social trait space.

Encoding and decoding models corroborate the RSA results and further reveal the critical social traits for the trait space

We next constructed encoding and decoding models to investigate the relationship between neural response and each individual social trait.

Using an encoding model (see Supplementary Methods), we identified subsets of neurons that significantly tracked social trait judgments (Pearson correlation between the mean normalized firing rate and the mean *z*-scored social trait ratings across 50 identities; see Fig. 2a for single-neuron examples for each social trait [Bonferroni correction for multiple comparisons] and Supplementary Fig. 7a for a summary of the number of significant neurons for each trait). This result revealed that neurons encoded the axes of the social trait space, suggesting a neuronal population code for social trait representations and corroborating the RSA results.

At the population level, we found that face-responsive neurons significantly encoded the judgments on the social traits associated with three comprehensive dimensions (warmth, competence, and femininity dimensions: warm, critical, competent, practical, feminine, and strong; Fig. 2b; two-tailed one-sample *t*-test of the correlation coefficient *r* against 0; practical and strong survived Bonferroni correction for multiple comparisons); and we observed similar results in all neurons (Supplementary Fig. 7b), all amygdala neurons (Supplementary Fig. 7c), and all hippocampal neurons (Supplementary Fig. 7d). Encoding of the social traits associated with the fourth comprehensive dimension (youth) was uniquely observed for the neural population in the amygdala (for youthful across face identities; Supplementary Fig. 7c). Our control analysis further confirmed our results (Supplementary Results).

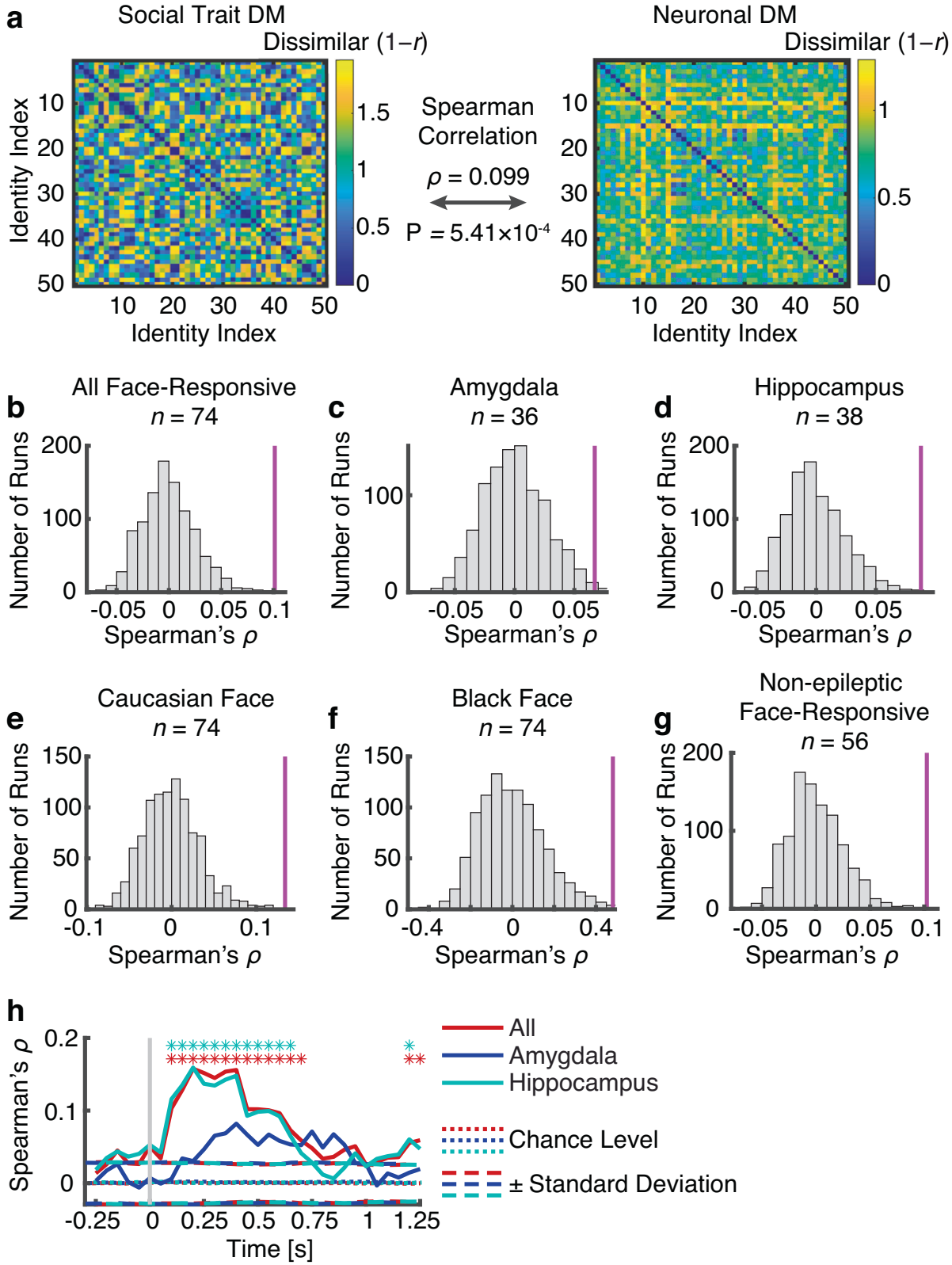
Notably, we explored whether different social traits were encoded with a similar latency. To answer this question, we investigated the temporal dynamics of encoding models using a moving window. We found that the social trait associated with the femininity dimension regarding gender (strong) was encoded earlier after stimulus onset than the social traits associated with the warmth and competence dimensions (warm, critical, and practical) which describe the more abstract personality characteristics of an individual (Fig. 2c) as opposed to physical characteristics. Therefore, our results indicate that different social trait dimensions may be processed at different stages in the brain with physical characteristics being processed earlier than more complex personality traits. Furthermore, although amygdala and hippocampal neurons showed similar encoding for most of the social traits (Supplementary Fig. 7c, d, g, h; see Supplementary Fig. 8 for temporal dynamics), we found that only amygdala neurons encoded the social trait youthful. Lastly, we found that the anterior and posterior hippocampus showed a similar encoding of social traits (Supplementary Fig. 7i, j).

Using a decoding model (see Supplementary Methods), we found that the neural population could predict the social trait judgments associated with all four dimensions (including the traits warm, critical, practical, feminine, strong, youthful, and charismatic) across face identities (Fig. 2d; Bonferroni correction for multiple comparisons) or across face images. Furthermore, we found similar results using partial least squares regression (Supplementary Fig. 7k) and regression with principal component analysis of neural responses (Supplementary Fig. 7l).

Together, the encoding and decoding models revealed that neurons encoded the axes of the social trait space and corroborated our finding that neurons from the amygdala and hippocampus collectively encode a comprehensive social trait space.

Neuronal social trait representation is general for different face stimuli and social trait spaces

We conducted an additional experiment to (1) rule out the possibility that participants' knowledge of some of the celebrities in our stimuli may influence neural representations of social traits, (2) investigate whether encoding of the social trait space can be



generalized to different face stimuli and a social trait space constructed using a different set of social traits, and (3) explore whether encoding of the social trait space is independent of the evaluative context.

We recorded from a separate population of 938 neurons (28 sessions from 8 patients; firing rate >0.15 Hz) while patients performed a trustworthiness judgment task (14 sessions; Fig. 3a) or a dominance judgment task (14 sessions) using the FaceGen

Fig. 1 A neuronal social trait space. a Correlation between dissimilarity matrices (DMs). The social trait DM (left matrix) was correlated with the neural response DM (right matrix). Color coding shows dissimilarity values ($1 - r$). **b**–**g** Observed vs. permuted correlation coefficient between DMs. The correspondence between DMs was assessed using permutation tests with 1000 runs. The magenta line indicates the observed correlation coefficient between DMs. The null distribution of correlation coefficients (shown in gray histogram) was calculated by permutation tests of shuffling the face identities. **b** All face-responsive neurons ($n = 74$). **c** Amygdala face-responsive neurons ($n = 36$). **d** Hippocampal face-responsive neurons ($n = 38$). **e** Social trait space constructed using Caucasian faces only ($n = 74$). **f** Social trait space constructed using Black faces only ($n = 74$). **g** Face-responsive neurons from the non-epileptic brain regions ($n = 56$). **h** Temporal dynamics of correlation between DMs. Bin size is 500 ms and step size is 50 ms. The first bin is from −500 to 0 ms (bin center: −250 ms) relative to stimulus onset, and the last bin is from 1000 to 1500 ms (bin center: 1250 ms) after stimulus onset. Dotted horizontal lines indicate the chance level and dashed horizontal lines indicate the ±standard deviation (SD) of the null distribution. The top asterisks illustrate the time points with a significant correlation between DMs (permutation test against null distribution, $p < 0.05$, corrected by false discovery rate (FDR) $Q < 0.05$).

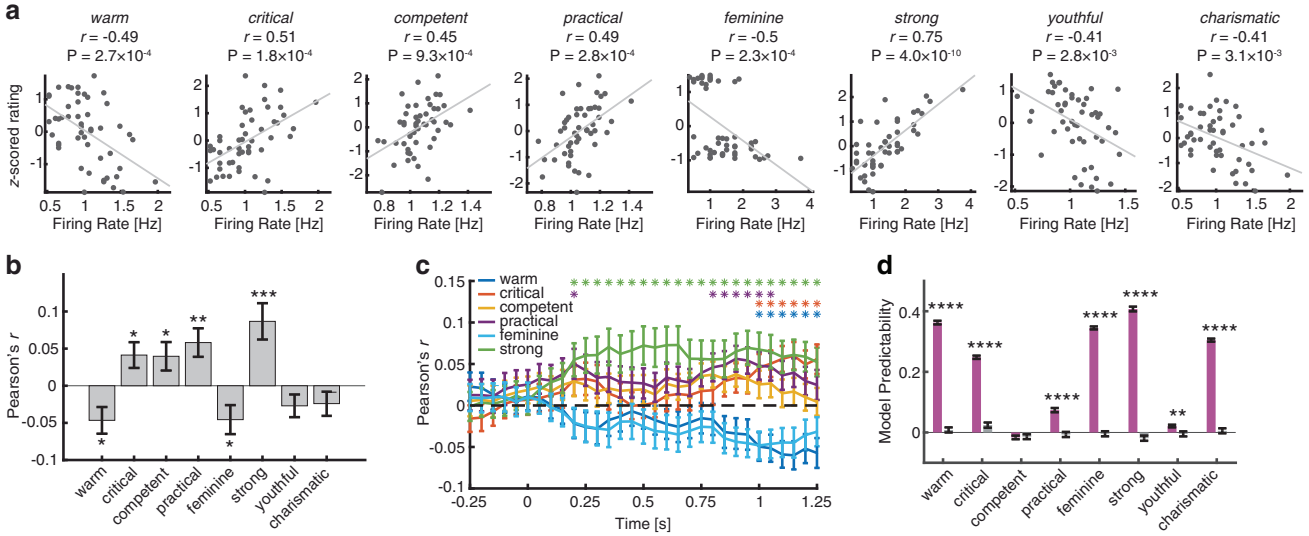


Fig. 2 Encoding and decoding models. a Example neurons that showed a significant correlation between the mean normalized firing rate and the mean *z*-scored rating for each social trait. Each dot represents a face identity, and the gray line denotes the linear fit. **b** Encoding of each social trait. The bars show the average correlation coefficient across all face-responsive neurons for each social trait. Error bars denote ±SEM across neurons. Asterisks indicate a significant difference from 0 (two-tailed paired *t*-test). $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$. **c** Encoding of different social traits over time. Error bars denote ±SEM across neurons. Asterisks shown on the top indicate a significant difference from 0 (two-tailed paired *t*-test, $p < 0.05$, corrected by false discovery rate (FDR) $Q < 0.05$). **d** Decoding of each social trait using a linear decoding model on face identities. Model predictability was assessed using the Pearson correlation between the predicted and actual trait ratings in the test dataset. The magenta bars show the observed response and the gray bars show the permuted response. Error bars denote ±SEM across permutation runs. Asterisks indicate a significant decoding performance (two-tailed two-sample *t*-test between observed vs. permuted). $**p < 0.01$ and $****p < 0.0001$.

model faces [12], which contained only feature information but no real identity information (Fig. 3a, b). We used nine social traits (attractiveness, competence, trustworthiness, dominance, mean, frightening, extroversion, threatening, and likability) to construct a social trait space (Fig. 3b). Again, we confirmed that the ratings from our patients were consistent with the consensus ratings from [12] (Pearson correlation: $r = 0.21 \pm 0.13$ (mean ± SD across sessions) for trustworthiness and $r = 0.34 \pm 0.17$ for dominance; two-tailed *t*-test against 0: both $ps < 0.05$).

Similarly, we calculated DMs between faces for social traits (Fig. 3c; using *z*-scored ratings for the nine social traits) and neural responses (Fig. 3d–f; using the mean normalized firing rate of neurons), and we assessed the correspondence between the social trait DM and the neural response DM using RSA. We found that the neural response DM was significantly correlated with the social trait DM (Fig. 3d, g; permutation $p = 0.010$), suggesting encoding of the social trait space was independent of face familiarity, specific face stimuli (natural photos of real people vs. computer-generated model faces), and specific social traits to construct the space. We also found that neurons encoded the social trait space separately in both the trustworthiness judgment task (Fig. 3e, h; permutation $p = 0.008$) and the dominance judgment task (Fig. 3f, i; permutation $p = 0.076$), suggesting that encoding of the social trait space could be independent of the evaluative context. Furthermore, in contrast to the natural photos of real people that may have emotional expressions, FaceGen model faces were all emotionally neutral, suggesting that amygdala and hippocampal neurons did not encode the social trait space based on emotion. Lastly, we separately analyzed amygdala and hippocampal neurons (Supplementary Results; Supplementary Figs. 9 and 10).

Together, this additional experiment not only confirmed our finding that neurons in the human amygdala and hippocampus encode a social trait space but also suggest that the neuronal social trait representation is general for different face stimuli and social trait spaces.

People with autism spectrum disorder (ASD) show an altered social trait representation

People with ASD demonstrate abnormal processing of social information from faces [15]. A specific neural structure hypothesized to underlie deficits in face processing in ASD is the amygdala, a brain structure that has long been implicated in autism [22, 31]. Therefore, in the present study we also explored whether people with ASD have a different social trait representation compared to controls and whether consensus ratings from participants with ASD can also predict neural responses from the amygdala and hippocampus.

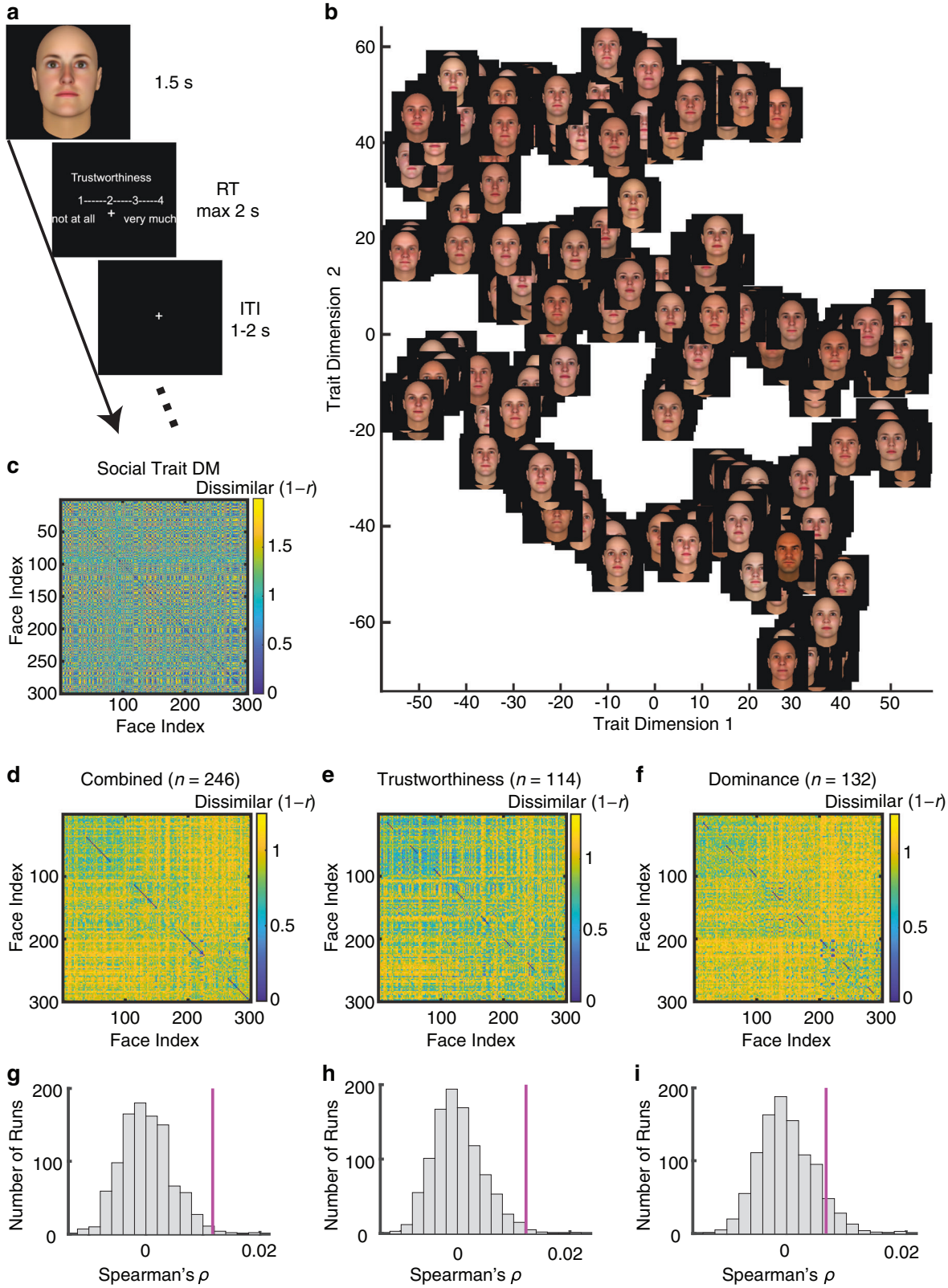


Fig. 3 Replication of the neuronal social trait space using the FaceGen stimuli and a different set of social traits. a Task. Each face was presented for 1.5 s, followed by participants' judgment of trustworthiness/dominance within 2 s. The inter-trial-interval (ITI) was jittered between 1 to 2 s. **b** Distribution of face images in the social trait space based on their consensus social trait ratings after dimension reduction using t-distributed stochastic neighbor embedding (t-SNE). **c** The social trait dissimilarity matrix (DM). **d**–**f** The neural response DMs for face-responsive neurons. Color coding shows dissimilarity values ($1 - r$). **g**–**i** Observed vs. permuted correlation coefficient between DMs. The magenta line indicates the observed correlation coefficient between DMs. The null distribution of correlation coefficients (shown in gray histogram) was calculated by permutation tests of shuffling the faces (1000 runs). **d**, **g** Combined trustworthiness and dominance judgment tasks. **e**, **h** Trustworthiness judgment task. **f**, **i** Dominance judgment task.

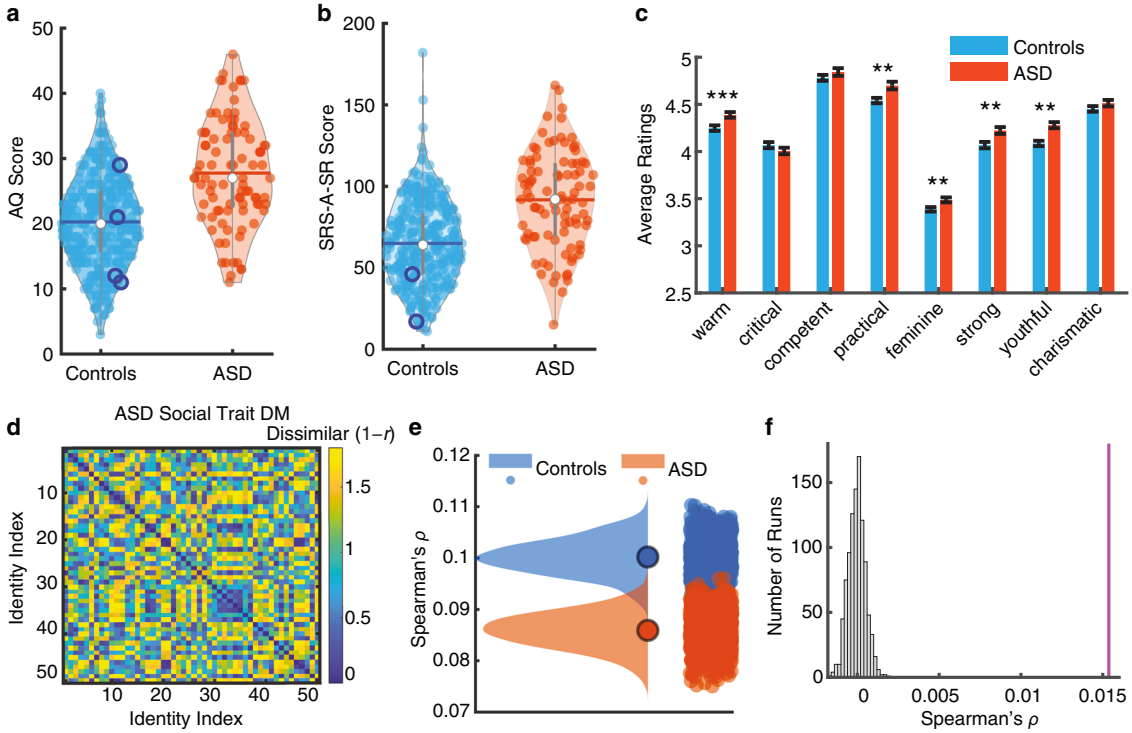


Fig. 4 People with ASD demonstrate atypical social trait representations. a Autism Spectrum Quotient (AQ). **b** Social Responsiveness Scale-2 Adult Self Report (SRS-A-SR). Violin plots present the median value as the white circle and the interquartile range as the gray vertical bars. Blue circles show the scores from neurosurgical patients, which are comparable to those from controls. **c** Social judgment rating for each trait. Error bars denote ±SEM across rating modules. Asterisks indicate a significant difference between participants with ASD and controls using two-tailed paired *t*-test. $^{**}p < 0.01$ and $^{***}p < 0.001$. **d** The social trait dissimilarity matrix (DM) from participants with ASD. **e** Bootstrap distribution of DM correspondence for each participant group. Blue: controls. Red: ASD. The dots show the mean value of each distribution. Participants with ASD showed a weaker correspondence with the neural response DM compared to controls. **f** Observed vs. permuted difference in DM correspondence between participant groups. The magenta line indicates the observed difference in DM correspondence between participant groups. The null distribution of difference in DM correspondence (shown in gray histogram) was calculated by permutation tests of shuffling the participant labels (1000 runs).

To address these questions, we acquired ratings of the CelebA stimuli from a sample of online participants with ASD (self-identified). We first confirmed that online participants with ASD demonstrated significantly higher scores compared to controls on standardized tests that evaluate ASD characteristics including the AQ (Fig. 4a; ASD: 27.76 ± 8.09 (mean ± SD), controls: 20.28 ± 6.82; two-tailed two-sample *t*-test: $t(427) = 8.94$, $p < 10^{-16}$) and SRS-A-SR (Fig. 4b; ASD: 91.73 ± 29.66, controls: 65.17 ± 25.19; $t(427) = 8.61$, $p = 1.11 \times 10^{-16}$). In comparison, neurosurgical patients had scores comparable to controls for both AQ (Fig. 4a; 18.25 ± 8.46; $t(339) = 0.59$, $p = 0.56$) and SRS-A-SR (Fig. 4b; 31.5 ± 20.51; $t(337) = 1.88$, $p = 0.06$). We also confirmed that online participants had scores similar to well-characterized in-lab participants from our prior ASD study [19] for both AQ (ASD: $t(108) = 0.73$, $p = 0.47$; controls: $t(349) = 1.42$, $p = 0.16$) and SRS-A-SR (ASD: $t(107) = 0.94$, $p = 0.35$; controls: $t(343) = 1.57$, $p = 0.16$).

We found that the social trait ratings differed in all four comprehensive dimensions (including the traits warm, practical, feminine, strong, and youthful) between participants with ASD and controls (Fig. 4c). Notably, although the social trait DM for participants with ASD (Fig. 4d) was similar to controls, it was less correlated with the neural response DM from the neurosurgical patients (derived with face-responsive neurons; $\rho = 0.084$ for ASD and $\rho = 0.10$ for controls; similar results were derived with all neurons). We used a bootstrapping approach to estimate the distribution of DM correspondence for each participant group (see "Methods") and we found that the two distributions were largely separated (Fig. 4e; the mean of the ASD distribution was significantly outside the control distribution ($p < 0.001$) and the mean of the control distribution was also significantly outside the ASD distribution ($p < 0.001$)). We also used a permutation test (see "Methods") and statistically confirmed that the difference in DM correspondence between participant groups was above chance (Fig. 4f; $p < 0.001$).

Together, although we did not directly acquire neural responses from participants with ASD, we found that the consensus ratings from the ASD group were less explainable of the neuronal responses in the amygdala and hippocampus. Our results thus indicate that the neuronal social trait space in the amygdala and hippocampus may have a behavioral consequence for social judgment and may account for abnormal social trait judgments of faces in ASD.

DISCUSSION

Our present results represent the first step toward constructing a social trait space for face processing at the single-neuron level in the human brain. We not only showed that single neurons encoded individual social traits when judging photos of famous people, but also demonstrated that the neuronal population in the amygdala and hippocampus encoded a comprehensive social trait space. In other words, we found that neurons encoded not only the axes composing the social trait space, but also the geometry of the space (i.e., pairwise distance between face images). In addition, we had a direct replication of our results using unfamiliar faces and different social traits, and our results further suggested that the neuronal social trait space in the amygdala and hippocampus may relate to abnormal social perception in autism based on behavioral data on social traits of the same stimuli.

Possible caveats

Our findings were based on recordings from neurosurgical epilepsy patients, so our study has the following clinical limitations. First, our patients may have different types of drug-resistant epilepsies and clinical variables (e.g., age of epilepsy onset, frequency of seizers, etc; Supplementary Table 1), which may confound our results. A future study with a larger sample of patients is needed to understand the impact of these epilepsy clinical variables on social trait judgments (e.g., using mixed-effects models adjusting these clinical variables, and/or analyzing social traits as a function of each clinical variable [correlation with a clinical variable or stratified by groups]). Second, our patients may have different levels of neuronal loss and gliosis as well as hippocampus sclerosis following mesial temporal lobe epilepsy, which may impact amygdala and hippocampus functions and interpretation of our results [32–35]. Although we did not measure neuronal loss, gliosis, or hippocampus sclerosis, it is worth noting that one of our patients (P7) had a prior right temporal lobectomy (primarily the right amygdala) but still demonstrated normal social trait judgments. Third, given the clinical restrictions on the locations of our neuronal sampling, our present study could only speak to the social trait coding in the amygdala and hippocampus, but other areas of the social brain [36] may also be associated with social trait judgment of faces (see [37] for a meta-analysis).

It is worth noting that our one-back task did not require patients to make any explicit face judgment (they simply indicated when a face was repeated); therefore, our analyses were relating neural responses of implicit face impressions provided by patients to the consensus ratings of explicit face impressions provided by an independent sample of over 400 participants from the general population. Using an additional experiment with computer-generated, unfamiliar faces, we further illustrated that encoding of the social trait space was independent of face familiarity, the knowledge of the face identity, as well as specific faces and traits being evaluated. Therefore, the neural coding in the amygdala and hippocampus can be a general mechanism for face evaluation and first impressions. Furthermore, because we used natural photos of real people as stimuli, some of the faces have emotional expressions (primarily happiness), which may bias social trait judgments [11]. However, our results with the emotionally neutral FaceGen model faces suggested that amygdala and hippocampal neurons did not encode the social trait space based on emotion.

The neural basis of social trait judgment

Past behavioral research has provided candidate dimensions for describing trait judgments of faces [12–14]; however, the biological bases of those psychological dimensions remain unknown. Here we showed that these dimensions were encoded by the neural population in the amygdala and hippocampus. We further showed that the neural correlates for different social trait dimensions varied in temporal dynamics (i.e., the femininity dimension [a physical characteristic] was encoded faster than the more abstract dimensions of warmth and competence [personality traits]), indicating that different categories of social trait information may arrive at the amygdala and hippocampus through different routes and thus at different latencies. This result is consistent with the notion that the amygdala connects with other parts of the brain through multiple routes [38]. In addition, we found that the dimensions of warmth, competence, and femininity were encoded across both face images and face identities, whereas the dimension of youth was only encoded across face identities regarding youthful and across face images regarding charismatic, likely because different face images from the same identity were more heterogenous along the youth dimension. Lastly, similar neural pattern analyses have been used to study race bias of faces using functional neuroimaging data [29] and face representation using intracranial electroencephalogram data [39] in humans. Similar representations of social traits may also be found elsewhere in the brain.

We found that both amygdala and hippocampal neurons encoded the social trait space (Fig. 1c, d and Supplementary Fig. 9), despite differences in response latency (Fig. 1h and Supplementary Fig. 9e) and specific traits encoded (Supplementary Figs. 7, 8, and 10). Although it has been suggested that the anterior and posterior hippocampus play different roles (i.e., the anterior hippocampus mediates anxiety-related behaviors whereas the posterior hippocampus is implicated in memory and spatial navigation [40]), we found that neurons from the anterior and posterior hippocampus had a similar representation and encoding of social traits (Supplementary Figs. 6 and 7i, j). A future study with more precise localization of the recording locations is needed to further elucidate the roles of different subregions of the amygdala and hippocampus in encoding social traits, especially given the revised views of the amygdala and hippocampal circuits [40, 41] and in the context of mesial temporal lobe epilepsy that may be associated with fear and anxiety [34, 35]. Lastly, it is worth noting that our present study only allowed us to establish a correlational relationship between neuronal activity and social trait judgments. A future study using neuromodulation (e.g., transitory treatment by deep brain stimulation, focused ultrasound) will be needed to further establish the causal relationship.

Face typicality and social trait judgment

Several social traits such as attractiveness and trustworthiness have been shown to be related to face typicality [42–45]; and in particular, the amygdala tracks face typicality [44]. Therefore, our current findings may be partially accounted by face typicality. Notably, face typicality has been considered as a social trait in the construction of our comprehensive social trait space [14]. We have shown that judgments of face typicality load most strongly on the competence dimension [14], which may particularly influence our results regarding the competence dimension. A future study is needed to elucidate the role of face typicality in neural representation of social traits, and our findings have provided candidate hypotheses to be tested in future research focusing on face typicality.

CONCLUSIONS

In conclusion, we for the first time revealed a neural social trait space at the single-cell level in humans. Encoding a comprehensive social trait space provides the neural basis for rapid spontaneous impressions of faces on multiple trait dimensions. Our present results are in line with the notion that face representations are encoded over a broad and distributed population of neurons [46], which has been conclusively demonstrated in the non-human primate IT cortex [3]. Our results further shed light on how face processing evolves along the visual processing stream where the brain transforms from encoding low-level facial features in the higher visual cortex to complex social traits in the amygdala and hippocampus. Our results also support the idea that the amygdala and hippocampus are highly involved in social perception and evaluation [6, 47], which in turn supports their roles in coding socially relevant and salient stimuli [20].

DATA AVAILABILITY

All data are publicly available on OSF (https://osf.io/a4jn3/).

REFERENCES

1. Willis J, Todorov A. First impressions: making up your mind after a 100-Ms exposure to a face. Psychological Sci. 2006;17:592–8.
2. Todorov A, Olivola CY, Dotsch R, Mende-Siedlecki P. Social attributions from faces: determinants, consequences, accuracy, and functional significance. Annu Rev Psychol. 2015;66:519–45.

3. Chang L, Tsao DY. The code for facial identity in the primate brain. Cell. 2017;169:1013–28.
4. Freiwald WA, Tsao DY, Livingstone MS. A face feature space in the macaque temporal lobe. Nat Neurosci. 2009;12:1187–96.
5. Leopold DA, Bondar IV, Giese MA. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature. 2006;442:572–5.
6. Montagrin A, Saiote C, Schiller D. The social hippocampus. Hippocampus. 2018;28:672–9.
7. Rutishauser U, Mamelak AN, Adolphs R. The primate amygdala in social perception—insights from electrophysiological recordings and stimulation. Trends Neurosci. 2015;38:295–306.
8. Adolphs R, Tranel D, Damasio AR. The human amygdala in social judgment. Nature. 1998;393:470–4.
9. Todorov A, Baron SG, Oosterhof NN. Evaluating face trustworthiness: a model based approach. Soc Cogn Affect Neurosci. 2008;3:119–27.
10. Wang S, Yu R, Tyszka JM, Zhen S, Kovach C, Sun S, et al. The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. Nat Commun. 2017;8:14821.
11. Said CP, Sebe N, Todorov A. Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. Emotion. 2009;9:260–4.
12. Oosterhof NN, Todorov A. The functional basis of face evaluation. Proc Natl Acad Sci USA. 2008;105:11087–92.
13. Sutherland CAM, Oldmeadow JA, Santos IM, Towler J, Michael Burt D, Young AW. Social inferences from faces: Ambient images generate a three-dimensional model. Cognition. 2013;127:105–18.
14. Lin C, Keles U, Adolphs R. Four dimensions characterize attributions from faces using a representative set of English trait words. Nat Commun. 2021;12:5168.
15. Adolphs R, Sears L, Piven J. Abnormal processing of social information from faces in autism. J Cogn Neurosci. 2001;13:232–40.
16. Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch Gen Psychiatry. 2002;59:809–16.
17. Pelphrey KA, Sasson NJ, Reznick JS, Paul G, Goldman BD, Piven J. Visual scanning of faces in autism. J Autism Developmental Disord. 2002;32:249–61.
18. Dawson G, Webb SJ, McPartland J. Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. Developmental Neuropsychol. 2005;27:403–24.
19. Wang S, Adolphs R. Reduced specificity in emotion judgment in people with autism spectrum disorder. Neuropsychologia. 2017;99:286–95.
20. Wang S, Adolphs R. Social saliency. In: Zhao Q, editor. Computational and cognitive neuroscience of vision. Singapore: Springer; 2017. p. 171–93.
21. Baron-Cohen S, Ring HA, Bullmore ET, Wheelwright S, Ashwin C, Williams SC. The amygdala theory of autism. Neurosci Biobehav Rev. 2000;24:355–64.
22. Schumann CM, Amaral DG. Stereological analysis of amygdala neuron number in autism. J Neurosci. 2006;26:7674–9.
23. Rutishauser U, Tudusciuc O, Wang S, Mamelak AN, Ross IB, Adolphs R. Single-neuron correlates of atypical face processing in autism. Neuron. 2013;80:887–99.
24. Kliemann D, Dziobek I, Hatri A, Baudewig J, Heekeren HR. The role of the amygdala in atypical gaze on emotional faces in autism spectrum disorders. J Neurosci. 2012;32:9469–76.
25. Freiwald W, Duchaine B, Yovel G. Face processing systems: from neurons to real-world social perception. Annu Rev Neurosci. 2016;39:325–46.
26. Cao R, Li X, Todorov A, Wang S. A flexible neural representation of faces in the human brain. Cereb Cortex Commun. 2020;1:tgaa055.
27. Reber TP, Bausch M, Mackay S, Boström J, Elger CE, Mormann F. Representation of abstract semantic knowledge in populations of human single neurons in the medial temporal lobe. PLoS Biol. 2019;17:e3000290.
28. Kriegeskorte N, Mur M, Bandettini P. Representational similarity analysis—connecting the branches of systems neuroscience. Front Syst Neurosci. 2008;2:4.
29. Stolier RM, Freeman JB. Neural pattern similarity reveals the inherent intersection of social categories. Nat Neurosci. 2016;19:795–7.
30. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57:289–300.
31. Baron-Cohen S, Ring HA, Bullmore ET, Wheelwright S, Ashwin C, Williams SCR. The amygdala theory of autism. Neurosci Biobehav Rev. 2000;24:355–64.
32. Yilmazer-Hanke DM, Wolf HK, Schramm J, Elger CE, Wiestler OD, Blümcke I. Subregional pathology of the amygdala complex and entorhinal region in surgical specimens from patients with pharmacoresistant temporal lobe epilepsy. J Neuropathol Exp Neurol. 2000;59:907–20.
33. Blümcke I, Thom M, Aronica E, Armstrong DD, Bartolomei F, Bernasconi A, et al. International consensus classification of hippocampal sclerosis in temporal lobe epilepsy: a Task Force report from the ILAE Commission on Diagnostic Methods. Epilepsia. 2013;54:1315–29.
34. Leal RB, Lopes MW, Formolo DA, de Carvalho CR, Hoeller AA, Latini A, et al. Amygdala levels of the GluA1 subunit of glutamate receptors and its phosphorylation state at serine 845 in the anterior hippocampus are biomarkers of ictal fear but not anxiety. Mol Psychiatry. 2020;25:655–65.
35. de Carvalho CR, Lopes MW, Constantino LC, Hoeller AA, de Melo HM, Guarnieri R, et al. The ERK phosphorylation levels in the amygdala predict anxiety symptoms in humans and MEK/ERK inhibition dissociates innate and learned defensive behaviors in rats. Mol Psychiatry. 2021;26:7257–69.
36. Stanley DA, Adolphs R. Toward a neural basis for social behavior. Neuron. 2013;80:816–26.
37. Mende-Siedlecki P, Said CP, Todorov A. The social evaluation of faces: a meta-analysis of functional neuroimaging studies. Soc Cogn Affect Neurosci. 2013;8:285–99.
38. Pessoa L, Adolphs R. Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. Nat Rev Neurosci. 2010;11:773–83.
39. Grossman S, Gaziv G, Yeagle EM, Harel M, Mégevand P, Groppe DM, et al. Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. Nat Commun. 2019;10:4934.
40. Strange BA, Witter MP, Lein ES, Moser EI. Functional organization of the hippocampal longitudinal axis. Nat Rev Neurosci. 2014;15:655–69.
41. Janak PH, Tye KM. From circuits to behaviour in the amygdala. Nature. 2015;517:284–92.
42. Langlois JH, Roggman LA. Attractive faces are only average. Psychological Sci. 1990;1:115–21.
43. Rhodes G. The evolutionary psychology of facial beauty. Annu Rev Psychol. 2005;57:199–226.
44. Todorov A, Mende-Siedlecki P, Dotsch R. Social judgments from faces. Curr Opin Neurobiol. 2013;23:373–80.
45. Sofer C, Dotsch R, Wigboldus DH, Todorov A. What is typical is good: the influence of face typicality on perceived trustworthiness. Psychological Sci. 2014;26:39–47.
46. Hinton GE, McClelland J, Rumelhart D. Distributed representations In Parallel distributed processing: explorations in the microstructure of cognition, eds. Rumelhart D, McClelland J, pp. 77–109. Cambridge, MA: MIT Press, 1986.
47. Adolphs R. What does the amygdala contribute to social cognition? Ann N Y Acad Sci. 2010;1191:42–61.

ACKNOWLEDGEMENTS

We thank all patients for their participation, staff from WVU Ruby Memorial Hospital for support with patient testing, and Paula Webster for valuable comments. This research was supported by AFOSR Young Investigator Program Award (FA9550-21-1-0088), NSF Grants (BCS-1945230, IIS-2114644), Dana Foundation Clinical Neuroscience Award, and ORAU Ralph E. Powe Junior Faculty Enhancement Award. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

RC, CL, XL, and SW designed research. RC and SW performed experiments. NJB performed surgery. RC, CL, JH, and SW analyzed data. RC, CL, XL, AT, and SW wrote the paper. All authors discussed the results and contributed toward the manuscript.

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/s41380-022-01583-x.

Correspondence and requests for materials should be addressed to Runnan Cao or Shuo Wang.