Schizophrenia Research: Cognition 45 (2026) 100431

Contents lists available at ScienceDirect

Schizophrenia Research: Cognition

journal homepage: www.elsevier.com/locate/scog

Internal representations of facial emotions in schizophrenia

Anita Song [a,b,*,1], Chengyu Zhang [b,1], Nicola Binetti [c], Sukhi S. Shergill [b,d], Isabelle Mareschal [e], Panayiota G. Michalopoulou [b]

[a] Birkbeck University of London, UK
[b] Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK
[c] University of Tor Vergata, Rome, Italy
[d] Kent and Medway Medical School, Canterbury, UK
[e] School of Biological and Behavioural Sciences, Queen Mary University of London, UK

ARTICLE INFO

Keywords:
Schizophrenia
Emotion processing
Emotion recognition
Social cognition
Internal representation
Facial expression
Genetic algorithms

ABSTRACT

Individuals with schizophrenia demonstrate atypical facial emotion recognition. However, inconsistent findings in the literature highlight the limitations of standard paradigms that rely on fixed, stereotypical facial configurations. The present study employed a novel computational tool to examine internal representations of facial emotions — defined as individual expectations of how emotions appear on the face. Twenty-eight patients with schizophrenia and 25 healthy controls generated facial expressions of happiness, fear, and anger on a photo-realistic avatar through an iterative selection process, converging on ideal expressions for each target emotion across 8 iterations of 10 samples per iteration. Individualised models capturing the range of facial configurations associated with each emotion category were constructed from the selected expressions. No significant group differences were observed in the number of expressions selected, the breadth of expressions deemed representative of each emotion (spread), or the discriminability of emotion categories (d-prime). Both groups demonstrated greater difficulty distinguishing fearful from angry expressions relative to distinguishing either from happy expressions. Notably, patients exhibited significantly greater within-group centroid dispersion, indicating that their internal representations were more variable and less similar compared to controls. This suggests that patients' internal representations of facial emotions are more heterogeneous, potentially reflecting less shared understanding of facial features that define each emotion category. These findings offer a novel, representation-based account of emotion recognition in schizophrenia. Rather than a uniform perceptual deficit, the results indicate greater variability in internal representations of facial expressions, which may disrupt emotion recognition and contribute to social communication difficulties.

1. Introduction

Traditional theories of emotion recognition propose that a small set of basic emotions—happiness, anger, sadness, fear, disgust, and surprise—are biologically determined and universally expressed because of their importance for social communication (Darwin, 1872; Ekman, 1992). However, this view has been challenged. Cross-cultural research has shown variability in how emotions are expressed and recognised (Elfenbein and Ambady, 2002; Jack et al., 2012), and neuroimaging studies suggest that brain activity during emotion perception aligns more strongly with conceptual knowledge about emotions than with specific visual features of faces (Brooks et al., 2019).

Together, these findings suggest that emotion recognition is shaped not only by visual input but also by a person's internal representation of how an emotion should look. We use the term internal representation to refer to an individual's mental model or expectation of the facial features associated with a particular emotion. According to theories of conceptual alignment, successful communication depends on people sharing sufficiently similar mappings between signals (e.g., facial expressions) and their meanings (Stolk et al., 2016). If individuals differ in their internal representations of emotions, this misalignment may lead to misunderstanding during social interactions. Therefore, studies that rely only on fixed, stereotypical facial expressions may overlook meaningful individual and group differences in how emotions are mentally

* Corresponding author at: Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK.
E-mail address: anita.song@kcl.ac.uk (A. Song).
[1] Joint first authorship.

https://doi.org/10.1016/j.scog.2026.100431
Received 6 January 2026; Received in revised form 15 March 2026; Accepted 15 March 2026

represented and interpreted (Barrett et al., 2019).

Patients with schizophrenia show well-documented difficulties in recognising facial emotions, which may contribute to social functioning impairments. A review found that patients perform worse than controls when identifying basic emotions, particularly negative emotions such as fear and anger (Kohler et al., 2010). Differences have also been observed in electrophysiological responses during multiple stages of facial emotion processing (Gao et al., 2021). However, most studies use highly stereotypical, posed expressions. This approach does not capture individual differences in how emotions are internally represented. Notably, David and Gibson (Davis and Gibson, 2000) found that when more natural ("genuine") expressions were used, patients with paranoid schizophrenia performed better than controls, reversing the typical impairment. This finding suggests that difficulties may partly reflect a mismatch between patients' internal representations and the stereotypical expressions typically used in experiments. Such mismatches may contribute to social miscommunication and reduced social reciprocity.

To better understand facial emotion processing in schizophrenia, research should move beyond forced-choice classification of standardised expressions and instead examine how individuals internally represent emotions. Recent studies have used computer-generated avatars to allow participants to iteratively select facial variants from sets of generated facial expressions, where selected faces were recombined and slightly modified across trials to converge on each participant's internal representation of a target emotion (Carlisi et al., 2021; Binetti et al., 2022). These studies demonstrate substantial variability across healthy individuals in the types of expressions associated with emotions such as happiness, anger, fear, and sadness. This variability suggests that differences in performance on traditional recognition tasks may reflect differences in internal representations rather than perceptual or affective deficits. Murray et al. (Murray et al., 2024) developed a method to model the distribution of expressions selected by each participant for a given emotion. This approach estimates (1) the average selected expression, (2) the spread or breadth of selected expressions, and (3) the discriminability between emotions. They found marked individual differences in these representation profiles. Moreover, individuals with more similar representation profiles tended to interpret facial expressions in similar ways, highlighting the potential relevance of internal representations for social communication.

The present study applied this approach to patients with schizophrenia and unaffected controls. Using a computer-generated avatar task (Binetti et al., 2022), participants selected facial expressions corresponding to three target emotions: happiness, anger, and fear. We estimated each participant's internal representation by modelling the distribution of expressions they selected for each emotion (Murray et al., 2024). We compared groups on: the number of expressions selected, the degree of dispersion in average selected expressions, the spread of selected expressions within each emotion, and the discriminability between emotions.

We hypothesised that patients might show differences in the spread and dispersion of their internal representations. Broader representations could reduce the specificity needed to distinguish emotions, whereas overly narrow representations might reduce sensitivity to variation. Greater dissimilarity between average representations for a given emotion could also lead to conceptual misalignment and contribute to communication difficulties. Finally, we explored whether alterations were more pronounced for negative emotions (anger and fear), given prior evidence of greater impairment for these emotions (Kohler et al., 2010). We also examined associations between internal representations and clinical characteristics, including symptom severity, illness duration, and antipsychotic medication use.

2. Methods

2.1. Participants

A total of 53 participants took part in the study, comprising 25 control participants and 28 patients with schizophrenia. Demographic details and group comparisons are shown in Table 1, and detailed information on participant ethnicity can be found in Table S1 of the Supplementary Materials. Compared to controls, patients had significantly lower years of education and IQ score, as measured by the Wechsler Abbreviated Scales of Intelligence II (WASI-II) two-subset form (Vocabulary and Matrix Reasoning subtests). We controlled for years of education in group comparisons, as IQ differences may be a core feature of schizophrenia pathology (McCutcheon et al., 2023).

Patients with schizophrenia were recruited from outpatient services of South London and Maudsley (SLAM) NHS Foundation Trust, and control participants were recruited through advertisements in the local community. All participants were 18–55 years old, right-handed, and scored less than 20 on the Beck Depression Inventory (Beck et al., 1961). Individuals with substance or alcohol dependence, systematic or neurological conditions, or treatment with CNS active medication other than antipsychotics were excluded from the sample. Control participants did not have a history of psychiatric illness or a first-degree relative currently or previously suffering from a psychotic illness. Patients had an ICD-10 diagnosis of schizophrenia or schizoaffective disorder, were treated with atypical antipsychotic monotherapy and were stable, with no worsening of symptoms or hospitalisation for at least 3 months prior to testing. Recruitment and testing were conducted at the Institute of Psychiatry, Psychology and Neuroscience, King's College London. All participants provided written informed consent and were compensated for their time and travel. Ethical approval was obtained by the London-Brighton & Sussex NHS National Research Ethics Committee (REC Ref Number22/LO/0358). The study was compliant with the Declaration of Helsinki.

2.2. Experimental stimuli

The experiment used a computer-based tool that allowed participants to gradually shape facial expressions displayed on a realistic 3D avatar by selecting expressions across iterative samples that match their idea of a target emotion. Facial expressions were generated by adjusting the geometry (vertices) of the avatar's face using combinations of higher-order parameters (blendshapes) (Roubtsova et al., 2021). These blendshapes approximate patterns of facial muscle movement, are based on the Facial Action Coding System (FACS) (Ekman and Friesen, 1978), and are used to compare performance.

Table 1
Participant demographics.

	Controls	Patients	Test statistic	*p*
Male/Female	11/14	21/7	chi-square	0.06
Age (years)	32.2(11.8)	37.9(8.4)	*t*-test	0.06
Education (years)	16.7(2.2)	14.0(2.9)	t-test	<0.001
WASI IQ	117.2 (12.8)	93.8(13.8)	t-test	<0.001
Duration of illness (years)	–	10.2(5.9)	–	–
CPZ equivalent	–	286.3 (105.5)	–	–
PANSS Positive	–	15.6(5.2)	–	–
PANSS Negative	–	12.3(4.8)	–	–
PANSS General	–	26.1(7.0)	–	–
PANSS Total	–	54.0(14.2)	–	–

2.3. Experimental procedure

Participants were instructed to generate facial expressions corresponding to three target emotions (happiness, fear, and anger), each completed separately using the computer tool. Expressions were displayed on a Caucasian male avatar (see Fig. 1).

Each trial began with 10 randomly generated facial expressions representing a wide range of facial configurations. On each trial, participants selected one or more expressions (between 1 and 10) that resembled the target emotion. Among these selections, they also identified the single expression that best matched the target emotion (elite face).

Selections were used to generate the next set of expressions. Features from the selected expressions were combined and augmented with random variation in a process that was based on evolutionary principles (Binetti et al., 2022), and the elite face always appeared in the next set, allowing efficient sampling of different facial configurations across trials. Over eight trials, this iterative selection procedure progressively converged toward the participant's preferred expression for the target emotion. The elite face selected on the final trial was taken as the participant's preferred expression for that emotion (final elite).

2.4. Analysis

2.4.1. Modelling expression representations

The output of the expression generation task consisted of weights on 149 blendshapes for each selected face. Of these, 41 core blendshapes corresponding to plausible FACS-based muscle movements were retained for analysis. Data were standardised (z-scored) prior to principal components analysis (PCA). PCA reduced the dimensionality from 41 blendshapes to 10 principal components, accounting for 36.29% of the variance. A list of the top 10 blendshapes ranked by PCA loading for the first two principal components is shown in Table S1 of the Supplementary Materials. The number of principal components was selected to balance variance explained and model stability. Models including more than 10 components failed to fit multiple participants' data, whereas using 10 components resulted in a poor fit for only one control participant. For each participant and each emotion, the distribution of all selected expressions across trials was estimated using multivariate Gaussian Kernel Density Estimation (KDE), following Murray et al. (Murray et al., 2024). This modelling approach provides a quantitative estimate of the participant's internal representation of the target emotion by characterising the distribution of selected facial configurations. All selected faces were weighted equally to avoid assumptions about the relative quality of selections. Scott's factor was used to define kernel bandwidth (Scott, 1992). Modelling was performed in Python 3.12.1.

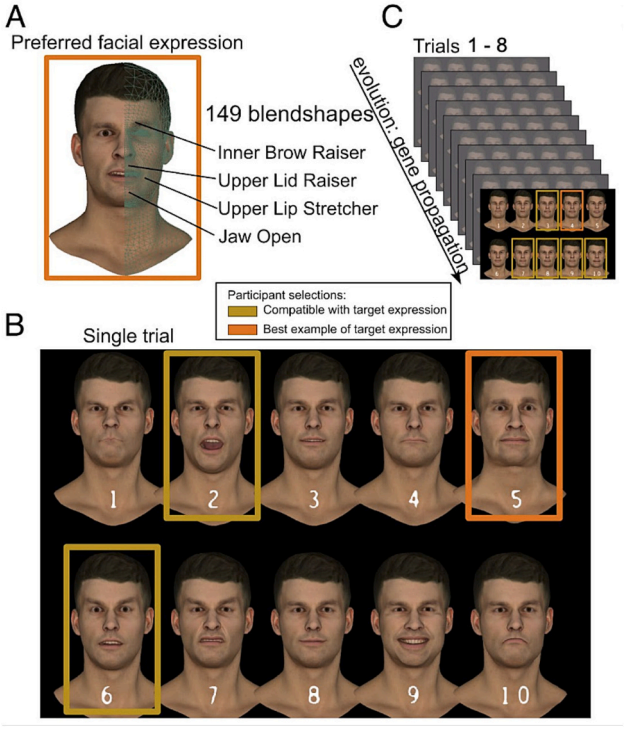

Fig. 1. Adapted from Binetti et al., (Binetti et al., 2022). **(A)** Avatar expressions were controlled using 149 facial expression parameters. **(B)** Within each trial, participants selected faces that matched the target emotion and identified the best-matching expression. **(C)** Across trials, new expressions were generated through an iterative selection procedure with random variation, converging toward the participant's preferred expression by the eighth trial.

2.4.2. Representation metrics

We derived several metrics from the task to characterise performance and internal representations. First, the total number of expressions selected across all trials for a given emotion was compared across groups. Then, for each participant and emotion, *centroid* was calculated as the average of the blendshape weights across all selected faces, representing the expression characteristics of a participant's average representation of that emotion. *Spread* was calculated as the trace (sum of diagonal elements) of the covariance matrix for each multivariate Gaussian KDE model, representing the breadth of facial configurations considered representative of a given emotion for an individual. Higher spread indicates a wider range of expressions considered representative of a given emotion. Fig. 2 shows a schematic illustration of centroid and spread. *D-prime* (d') was operationalised as [d-prime = abs(distance between centroids)/sqrt(mean(spreads))], representing the discriminability or separation between representations of different emotions. This was calculated for each pair of emotions (happy/fear, happy/anger, fear/anger). Lower d' values indicate greater overlap and reduced distinctiveness between emotion representations.

Centroid dispersion was defined as the average pairwise Euclidean distance between an individual's centroid and the centroids of other participants, representing the degree of dissimilarity between the individual's average representation and others' average representations. *Within-group centroid dispersion* refers to the centroid dispersion between a given participant and all other participants in the same group, and *between-group centroid dispersion* refers to the centroid dispersion between a given participant and all participants in the other group. Higher centroid dispersion suggests greater interindividual variability in average representations at the group level. Group differences in within-group centroid dispersion were examined. Between-group centroid dispersion was compared to within-group centroid dispersion separately for patients and controls to assess overlap in average representations at the group level.

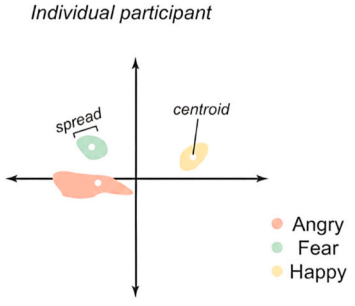

Fig. 2. Adapted from Murray et al., (Murray et al., 2024). Schematic illustration of an individual's emotion representations in the reduced feature space (from PCA), depicting centroid and spread. The separation between representations determines d-prime. Actual centroid locations are depicted in Fig. 5A in the results.

2.5. Statistics

Generalized Estimating Equations (GEE) were used to examine the main effects of group (patients, controls), emotion (happy, fearful, angry), and their interaction on each task metric, while adjusting for years of education:

$$\text{metric} \sim \text{group x emotion} + \text{education}$$

where the interaction term was not significant, results from the main-effects model are reported. An exchangeable working correlation structure was used to account for within-subject clustering (52 clusters; three observations per cluster). Analyses were implemented using the *geeglm* function from the *geepack* library (Halekoh et al., 2006). Omnibus effects were evaluated using Type III Wald tests. False discovery rate was controlled at 0.05 using the Benjamini–Hochberg procedure across four primary group comparisons (number of selections, within-group centroid dispersion, spread, and d-prime). GEE was also used to examine centroid dispersion as a function of distance type (within-group, between-group) and emotion for each group (patients and controls):

$$\text{centroid dispersion} \sim \text{distance type x emotion}$$

Post-hoc comparisons were conducted using estimated marginal means (EMMs) from the emmeans package (Lenth, 2023), with Tukey HSD adjustment. Associations between internal representation metrics and clinical measures (symptom severity, duration of illness, antipsychotic medication, and cognitive flexibility) were examined using multiple regression models in R (e.g., *symptom severity* ~ *angry centroid dispersion* + *fearful centroid dispersion* + *happy centroid dispersion*) (Chambers, 1992). All statistical analyses were conducted in R version 4.5.2.

3. Results

Participants used a computational tool to iteratively generate facial expressions corresponding to their internal representations of anger, fear, and happiness. On each trial, they selected up to 10 expressions that reflected the target emotion, across a total of eight trials. A multivariate Gaussian Kernel Density Estimation (gKDE) model was fit to all faces selected by each individual for a given emotion. This model estimates the distribution of selected expressions and provides a quantitative characterisation of each participant's internal representation—that is, the range of facial configurations they considered representative of the target emotion. We were unable to fit the model to one participant's data; this participant was excluded from further analyses, resulting in a final sample of $N = 52$ (24 controls, 28 patients). Fig. 3 illustrates the within-group average of the best-matching expression selected on the final trial, taken as the participant's ideal expression for that emotion.

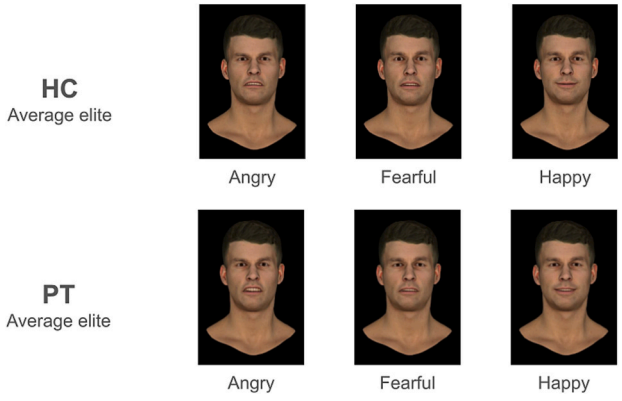

Fig. 3. Averaged final elites of each target emotion for controls (top) and patients (bottom), rendered onto the same Caucasian male avatar that was used during the experiment.

3.1. Number of selections

We compared the total number of selections made by patients and controls across all trials for each emotion (happy, angry, fearful, Fig. 4A). The group × emotion interaction was not statistically significant ($p = .64$). The main effect of group on the total number of faces selected was not significant ($F(1) = 3.05$, $p_{adj} = 0.16$). There was a significant main effect of target emotion on the number of selections ($F(2) = 8.75$, $p_{adj} < 0.001$). Across groups, the effect of emotion on the number of selections was driven by fewer selections for fearful faces compared to angry faces ($p < .001$). Other pairwise comparisons between emotions were nonsignificant ($p > .1$).

The effect of group on the number of selections was also not significant when controlling for trial number ($F(1) = 3.1$, $p = .08$). Both groups made fewer selections during earlier trials than later trials, as expected when using the iterative procedure (Fig. 4B). We also calculated the distance between *selected* faces and the centroid, and found that it decreased across trials, while the distance between *nonselected* faces and the centroid was consistently higher across trials (Fig. S1 in the Supplementary Materials). We found the same pattern for controls, which suggests that patients engaged with the experimental tool in a consistent and goal-directed manner (similar to controls), converging on an ideal expression by the final trial.

3.2. Within-group and between-group variability

Centroids representing the average of all selected faces for each emotion are shown in Fig. 5A. To characterise variability in the position of these average expressions, we calculated two centroid dispersion measures for each participant: (1) the average pairwise Euclidean distance between their centroid and the centroids of all other participants in the same group (within-group dispersion; Fig. 5B), and (2) the average pairwise Euclidean distance between their centroid and the centroids of participants in the other group (between-group dispersion).

First, we examined the effects of group (patients vs. controls) and emotion (happy, angry, fearful) on within-group dispersion. The group × emotion interaction was not significant ($p = .15$). There was a significant main effect of group on within-group dispersion ($F(1) = 19.7$, $p_{adj} < 0.001$), with patients showing higher within-group dispersion than controls. Emotion also had a significant main effect ($F(2) = 25.85$, $p_{adj} < 0.001$). Across groups, within-group dispersion was greatest for fear, followed by anger, and was lowest for happiness. All pairwise comparisons between emotions were significant (all p's < 0.001).

Next, we examined the effects of distance type (within-group vs. between-group) and emotion on centroid dispersion. For controls, there was a significant interaction between distance type and emotion ($F(2) = 13.2$, $p < .001$). Within-group dispersion was consistently lower than between-group dispersion across all three emotions (all p's < 0.001), indicating that controls' average representations were more similar to one another than to those of patients. In patients, the interaction was also significant ($F(2) = 16.1$, $p < .001$), but the pattern differed: within-group dispersion was higher than between-group dispersion across all emotions (all p's < 0.01). This indicates that patients showed greater variability among themselves than relative to controls. Overall, centroid dispersion was greatest within the patient group, intermediate between patients and controls, and lowest within the control group (Table 2).

3.3. Spread and discriminability in emotion representations

The spread of an individual's internal representation was defined as the sum of the diagonal elements of the covariance matrix of the fitted gKDE for each emotion (Fig. 6A). Spread therefore reflects the range of facial expressions an individual considered representative of a given

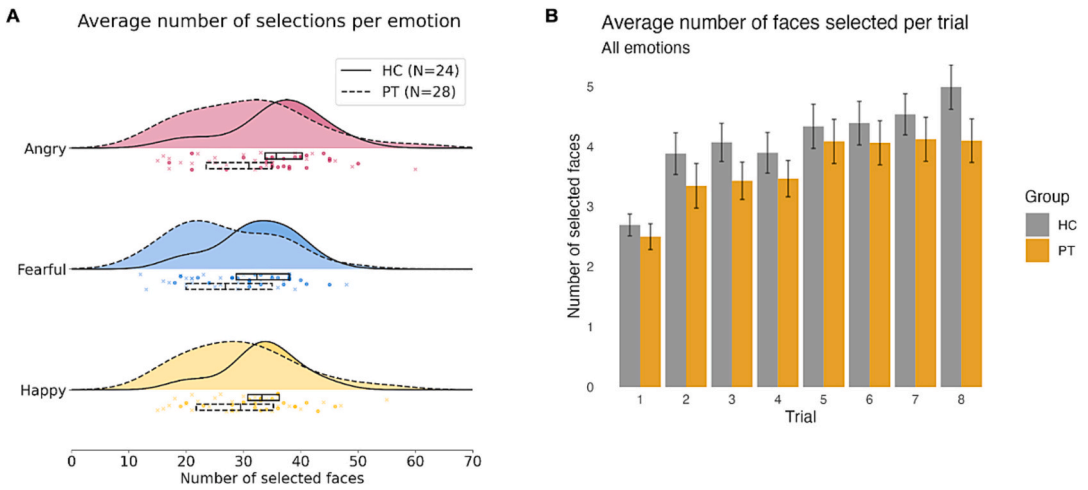

Fig. 4. The number of faces selected by each participant in the experimental task. **(A)** Total number of selections for each emotion category, averaged across patients (dotted line), and controls (solid line). **(B)** Average number of selections by patients (yellow) compared to controls (grey) in each trial (1–8), across all emotions (angry, fearful, happy). Error bars represent 95% confidence intervals.

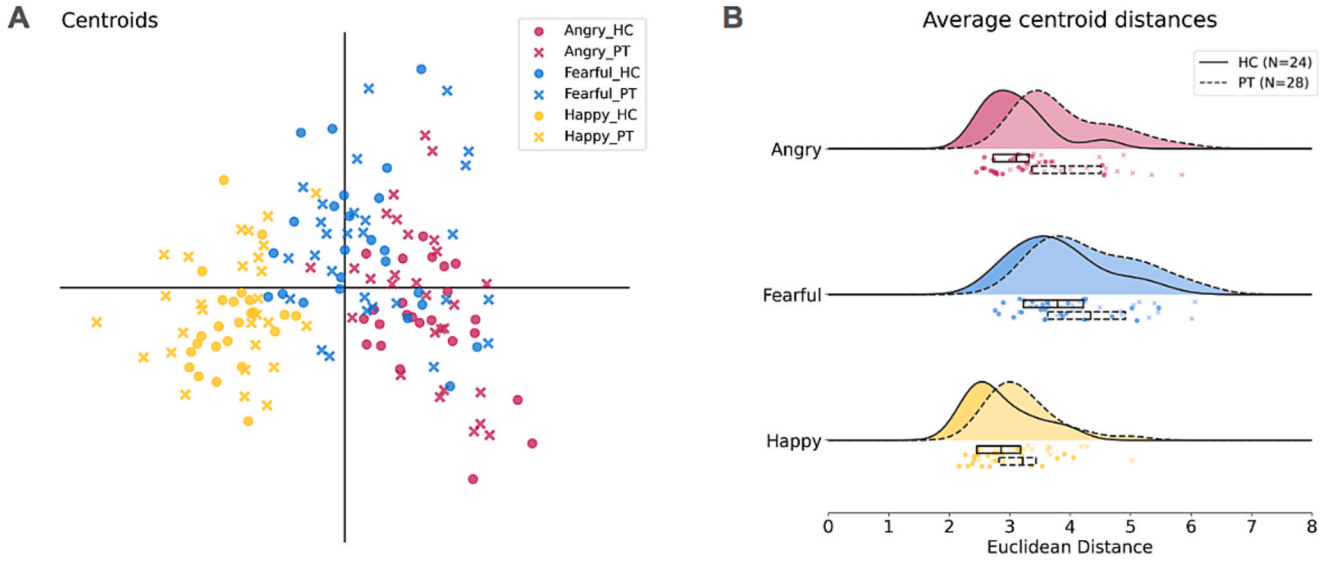

Fig. 5. Variability in participants' internal representations across groups. **(A)** Position of centroids (average selected expressions) plotted along the first two principal components of the reduced feature space. Each point represents an individual participant's centroid for a given emotion. **(B)** Within-group centroid dispersion. For each participant, the average pairwise Euclidean distance between their centroid and those of other participants in the same group was calculated for each emotion. Patients showed significantly greater within-group centroid dispersion than controls ($p = .001$).

Table 2
Average EMM of within- and between-group centroid dispersion across emotions.

	Within-group	Between-group
Controls	3.25 (0.07)	3.54 (0.07)
Patients	3.82 (0.08)	3.54 (0.10)

emotion. The group × emotion interaction was not statistically significant ($p = .93$). There was no main effect of group on spread ($F(1) = 0.19$, $p_{adj} = 0.66$), indicating that patients and controls did not differ in the overall size of their internal representations. There was a significant main effect of emotion ($F(2) = 23.25$, $p_{adj} < 0.001$). Across groups, spread was significantly lower for happiness compared to both anger and fear (all p's < 0.001), while anger and fear did not differ from one another ($p = .59$).

To assess the discriminability between emotions, we calculated d-prime (d') for each pairwise combination of representations (fearful/happy, angry/happy, angry/fearful; Fig. 6B). The group × emotion interaction was not statistically significant ($p = .86$). There was no main effect of group on d' ($F(1) = 0.98$, $p_{adj} = 0.32$), indicating comparable emotion discriminability in patients and controls. There was a significant main effect of emotion pair ($F(2) = 6.21$, $p_{adj} = 0.003$) on discriminability. The angry/fearful comparison yielded significantly lower d' values than both angry/happy and fearful/happy (all p's < 0.05), while angry/happy did not differ from fearful/happy ($p = .86$). A summary of the results from group comparisons models are provided in Table 3, and post-hoc estimated marginal means are reported in Table 4.

3.4. Associations with schizophrenia-related characteristics

We selectively explored the relationship between within-group centroid dispersion and schizophrenia-related variables in the patient group, to examine if they vary across pathological profiles given the

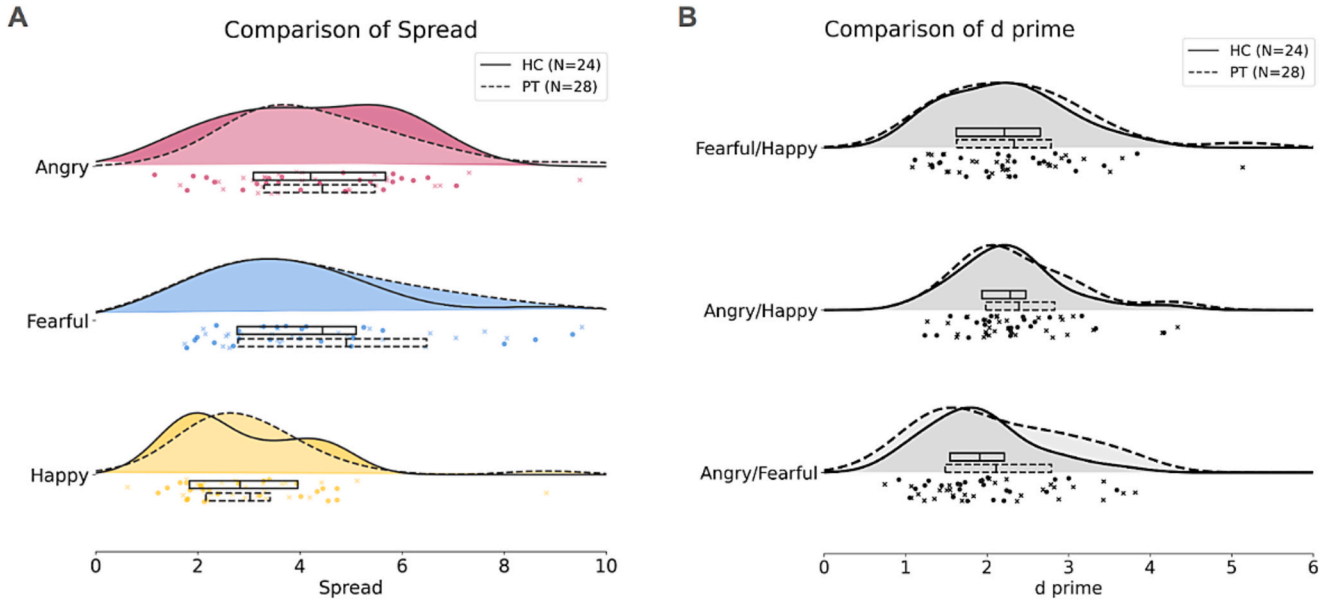

Fig. 6. We also looked for group differences in spread and discriminability of internal emotion representations. **(A)** Distribution of spreads for patients and controls (i.e. perceptive field size). **(B)** Distribution of d-primes for patients and controls for each emotion pair (fearful vs happy, angry vs happy, angry vs fearful).

Table 3
Summary of GEE results.

		df	*F*	*p*	p_{adj}
Number of selections	Group	1	3.05	0.083	0.16
	Emotion	2	8.75	<0.001	<0.001
Within-group centroid dispersion	Group	1	19.69	<0.001	**<0.001**[a]
	Emotion	2	25.85	<0.001	<0.001
Spread	Group	1	0.19	0.66	0.66
	Emotion	2	23.25	<0.001	<0.001
D-prime	Group	1	0.98	0.32	0.43
	Emotion	2	6.21	0.003	0.003

[a] Statistically significant group difference after adjusting for multiple comparisons.

Table 4
Summary of post-hoc EMM results.

	Patients	Controls	Emotions
N selections	29.6 (1.50)	33.2 (1.37)	Fearful: 29.5 (1.08) Angry: 33.3 (1.30) Happy: 31.4 (1.11)
Within-group centroid dispersion	**3.83 (0.09)**[a]	**3.24 (0.09)**[b]	Fearful: 4.06 (0.11) Angry: 3.52 (0.09) Happy: 3.03 (0.08)
Spread	4.07 (0.28)	3.89 (0.28)	Happy: 2.93 (0.18) Angry: 4.33 (0.23) Fearful: 4.69 (0.35)
D-prime	2.3 (0.14)	2.11 (0.11)	Angry/happy: 2.33 (0.09) Fearful/happy: 2.27 (0.11) Angry/fearful: 2.01 (0.10)

[a] Adjusted mean of patient group for metric that differed significantly between groups.
[b] Adjusted mean of control group for mnetric that differed significantly between groups.

significant difference between patients and controls. Using the PANSS total score as a measure of symptom severity, we found no relationship between symptom severity and within-group centroid dispersion ($R^2 = 0.02$, $p = .34$). Relationships to PANSS positive and negative subscales were also examined, revealing no significant associations. Duration of illness did not relate to within-group centroid dispersion ($R^2 = -0.005$, $p = .43$). Finally, we examined the medication use, estimated as the Chlorpromazine equivalent dosage. We found no association between medication use and within-group centroid dispersion ($R^2 = -0.06$, $p = .70$).

4. Discussion

This study examined internal representations of facial expressions in patients with schizophrenia and controls using participant-generated facial expressions corresponding to three target emotions (happy, angry, fearful). A computational modelling approach was used to estimate, for each participant and emotion, the central tendency (centroid), spread (within-individual range), and discriminability between emotion categories.

We found significantly greater centroid dispersion in patients than in controls. In contrast, groups did not differ in the number of selections, spread or discriminability between emotions. Moreover, representation metrics were not associated with symptom severity, illness duration, or antipsychotic medication dose.

4.1. Spread and discriminability of emotion representations

Patients did not differ from controls in the spread of their internal representations. That is, the range of facial expressions each individual considered representative of a given emotion was comparable across groups. Across both groups, spread was larger for angry and fearful expressions than for happy expressions. One interpretation is that negative emotions may be represented with broader acceptable variation, possibly reflecting heightened sensitivity to threat-related signals (Green and Phillips, 2004).

Similarly, patients and controls did not differ in discriminability between emotion categories. For both groups, angry and fearful representations were less discriminable from one another than either was from happy. This pattern is consistent with prior work in non-clinical samples using similar generative methods, suggesting that happiness is represented more distinctly from negative emotions (Binetti et al., 2022).

The absence of group differences in spread or discriminability was

somewhat unexpected given consistent reports of impaired recognition of negative emotions in schizophrenia (Kohler et al., 2010). However, emotion recognition deficits are heterogeneous and may vary across clinical subgroups (Fusar-Poli et al., 2022). The present findings suggest that, at least at the level of representational range and category separation, patients do not exhibit a uniformly broader or more overlapping internal structure of emotion categories.

4.2. Interindividual variability of emotion representations

In contrast to spread and discriminability, we observed robust group differences in centroid dispersion. Patients showed significantly greater within-group centroid dispersion than controls, indicating less consensus about the central features of each emotion category. Notably, centroid dispersion within patients exceeded both the centroid dispersion observed within controls and the centroid dispersion between patients and controls. This pattern suggests that disagreement in category-specific expression traits is especially pronounced within the patient group itself.

Importantly, this effect reflects between-person differences in where internal representations are centred, rather than broader representations within individuals. Because spread did not differ between groups, patients were not defining emotions more loosely or tightly; rather, they differed from one another more strongly in which expression features categorised each emotion. One possible account concerns altered face-processing strategies in schizophrenia. Previous research has reported reduced configural processing and greater reliance on individual facial features during emotion perception (Joshua and Rossell, 2009). If patients differentially weight specific facial features when constructing internal representations (e.g., focusing on the mouth versus the eyes), this could shift the location of their centroids without expanding representational spread. Although facial feature weighting was not directly investigated here, this mechanism provides a plausible explanation for increased centroid dispersion without changes in spread.

An alternative account is that centroid dispersion reflects greater heterogeneity in experience-dependent priors across patients. From a predictive processing perspective, internal representations are shaped by accumulated social experience (Goel and Gendron, 2025). Greater variability in social environments, symptom profiles, or socio-cognitive functioning could lead to more idiosyncratic expectations about how emotions are expressed.

Importantly, increased representational diversity is not inherently maladaptive. Greater within-individual breadth could, in principle, support flexible interpretation. However, the present pattern reflects disagreement across individuals about category-defining features, rather than broader representations within individuals. If facial expressions function as social signals, successful communication may depend partly on alignment between interlocutors' internal representations. Greater dispersion across individuals could therefore increase the likelihood that the same facial configuration is mapped onto different emotion categories, elevating the potential for miscommunication.

This interpretation aligns with the framework of conceptual alignment (Stolk et al., 2016), which proposes that effective social interaction depends on shared internal models. However, we did not directly measure communicative alignment or emotion recognition accuracy. Future studies could test whether dyads with more similar centroids show greater agreement in emotion judgments or more efficient communicative coordination, and whether centroid dispersion predicts real-world social functioning.

4.3. Schizophrenia-related characteristics

We did not observe associations between centroid dispersion and symptom severity, illness duration, or medication dose. This pattern is consistent with meta-analytic findings suggesting weak or inconsistent relationships between positive and negative symptom scales and emotion recognition performance (Keefe et al., 2006). One interpretation is that representational dispersion reflects a relatively trait-like characteristic rather than a state-dependent effect of symptom fluctuations.

However, clinical scales vary in their sensitivity to socio-affective processing, and our sample size may have limited power to detect small effects. Larger samples, ideally with dimensional assessments of paranoia, social cognition, and functioning, may clarify whether representational heterogeneity relates to specific symptom dimensions.

4.4. Limitation and future directions

Several limitations should be considered. First, the sample size was modest, which may have limited power to detect subtle interactions or clinical associations. Nonetheless, the observed group difference in centroid dispersion was statistically robust.

An additional consideration is the potential influence of other-race effects on emotion perception. Both healthy individuals and patients with schizophrenia typically recognise emotions more accurately on same-race faces than on other-race faces, with decreased discriminability for anger and other expressions when displayed on less familiar racial groups (Pinkham et al., 2008; Jiang et al., 2023). Using a single Caucasian male avatar may therefore limit the generalisability of the findings and could contribute to variability in internal representations if participants differ in experience with that facial phenotype. We conducted an additional analysis to examine the group difference in within-group centroid dispersion while controlling for race and sex, and found that this did not change the observed group effect (details reported in section 4 of the Supplementary Materials). Nonetheless, future studies should incorporate multiple avatars varying in race and sex to test whether internal emotion representations generalise across socially and culturally diverse faces and to ensure that observed differences are not confounded by race-related perceptual biases.

Finally, only three basic emotions were examined, and no neutral condition was included. Future work should test whether dispersion extends to additional emotion categories and how representations are positioned relative to neutrality. Fourth, we did not include a standard emotion recognition task. As such, we cannot directly link representational metrics to behavioural recognition accuracy. However, by isolating internal representations from speeded categorisation demands, the present design reduces confounds related to processing speed, attentional load, or decision strategies. This allows clearer examination of representational structure itself.

Future research should combine generative modelling approaches with behavioural emotion recognition tasks, socio-cognitive measures, and real-world functioning indices. Such integration would help determine whether increased interindividual dispersion contributes to emotion recognition variability and to broader social communication difficulties.

5. Conclusion

In summary, patients with schizophrenia did not differ from controls in the spread or discriminability of internal emotion representations. However, they exhibited greater differences in the central features of those representations. This pattern suggests not a uniformly broader or narrower definition of emotions, but reduced consensus about which facial cues define each emotion category. Such divergence in internal representations may increase alignment demands during social interaction and contribute to communication difficulties. These findings provide a mechanistic account of how emotion-processing differences in schizophrenia may arise from variability in representational structure rather than from global deficits in representational breadth or category separation.

CRediT authorship contribution statement

Anita Song: Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis, Data curation, Conceptualization. **Chengyu Zhang:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. **Nicola Binetti:** Writing – review & editing, Validation, Software, Formal analysis, Data curation, Conceptualization. **Sukhi S. Shergill:** Writing – review & editing, Visualization, Validation, Supervision, Formal analysis, Data curation, Conceptualization. **Isabelle Mareschal:** Writing – review & editing, Visualization, Validation, Supervision, Software, Formal analysis, Data curation, Conceptualization. **Panayiota G. Michalopoulou:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding sources

The study is funded by MRC - Clinical Academic Research Partnership (CARP) to Principal Investigator Dr. Panayiota Michalopoulou. The funder's reference is MR/V037218/1.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the participants who took part in the study, and Dr. Thomas Murray for his insights into the analysis strategy.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scog.2026.100431.

References

Barrett, L.F., Adolphs, R., Marsella, S., Martinez, A., Pollak, S.D., 2019. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. Psychol. Sci. Public Interest 20 (1), 1–68. https://doi.org/10.1177/1529100619832930.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. Arch. Gen. Psychiatry 4, 561–571.

Binetti, N., Roubtsova, N., Carlisi, C., Cosker, D., Viding, E., Mareschal, I., 2022. Genetic algorithms reveal profound individual differences in emotion recognition. Proc. Natl. Acad. Sci. USA 119 (45), e2201380119. https://doi.org/10.1073/pnas.2201380119.

Brooks, J.A., Chikazoe, J., Sadato, N., Freeman, J.B., 2019. The neural representation of facial-emotion categories reflects conceptual structure. Proc. Natl. Acad. Sci. USA 116 (32), 15861–15870. https://doi.org/10.1073/pnas.1816408116.

Carlisi, C.O., Reed, K., Helmink, F.G.L., Lachlan, R., Cosker, D.P., Viding, E., Mareschal, I., 2021. Using genetic algorithms to uncover individual differences in how humans represent facial emotion. R. Soc. Open Sci. 8 (10), 202251. https://doi.org/10.1098/rsos.202251.

Chambers, J.M., 1992. Linear models. In: Chambers, J.M., Hastie, T.J. (Eds.), Statistical Models in S. Wadsworth & Brooks/Cole, Pacific Grove (CA), pp. 95–144.

Darwin, C., 1872. The Expression of the Emotions in Man and Animals. Cambridge University Press, Cambridge.

Davis, P.J., Gibson, M.G., 2000. Recognition of posed and genuine facial expressions of emotion in paranoid and nonparanoid schizophrenia. J. Abnorm. Psychol. 109 (3), 445–450.

Ekman, P., 1992. An argument for basic emotions. Cognit. Emot. 6 (3–4), 169–200. https://doi.org/10.1080/02699939208411068.

Ekman, P., Friesen, W.V., 1978. Facial Action Coding System. Consulting Psychologists Press, Palo Alto.

Elfenbein, H.A., Ambady, N., 2002. On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychol. Bull. 128 (2), 203–235.

Fusar-Poli, L., Pries, L.K., van Os, J., et al., 2022. Examining facial emotion recognition as an intermediate phenotype for psychosis. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 113, 110440. https://doi.org/10.1016/j.pnpbp.2021.110440.

Gao, Z., Zhao, W., Liu, S., Liu, Z., Yang, C., Xu, Y., 2021. Facial emotion recognition in schizophrenia. Front. Psych. 12, 633717. https://doi.org/10.3389/fpsyt.2021.633717.

Goel, S., Gendron, M., 2025. Why internal representations and their temporal dependence matter for emotion inference. Emot. Rev. 18, 42–56.

Green, M.J., Phillips, M.L., 2004. Social threat perception and the evolution of paranoia. Neurosci. Biobehav. Rev. 28 (3), 333–342. https://doi.org/10.1016/j.neubiorev.2004.03.006.

Halekoh, U., Højsgaard, S., Yan, J., 2006. The R package geepack for generalized estimating equations. J. Stat. Softw. 15 (2).

Jack, R.E., Garrod, O.G.B., Yu, H., Caldara, R., Schyns, P.G., 2012. Facial expressions of emotion are not culturally universal. Proc. Natl. Acad. Sci. USA 109 (19), 7241–7244. https://doi.org/10.1073/pnas.1200155109.

Jiang, Z., Recio, G., Li, W., Zhu, P., He, J., Sommer, W., 2023. The other-race effect in facial expression processing: behavioral and ERP evidence from a balanced cross-cultural study in women. Int. J. Psychophysiol. 183, 53–60. https://doi.org/10.1016/j.ijpsycho.2022.11.009.

Joshua, N., Rossell, S., 2009. Configural face processing in schizophrenia. Schizophr. Res. 112 (1–3), 99–103. https://doi.org/10.1016/j.schres.2009.03.033.

Keefe, R.S.E., Bilder, R.M., Harvey, P.D., et al., 2006. Baseline neurocognitive deficits in the CATIE schizophrenia trial. Neuropsychopharmacology 31 (9), 2033–2046. https://doi.org/10.1038/sj.npp.1301072.

Kohler, C.G., Walker, J.B., Martin, E.A., Healey, K.M., Moberg, P.J., 2010. Facial emotion perception in schizophrenia: a meta-analytic review. Schizophr. Bull. 36 (5), 1009–1019. https://doi.org/10.1093/schbul/sbn192.

Lenth, R.V., 2023. emmeans: estimated marginal means, aka least-squares means. R package version 1.8.5.

McCutcheon, R.A., Keefe, R.S.E., McGuire, P.K., 2023. Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment. Mol. Psychiatry 28, 1902–1918. https://doi.org/10.1038/s41380-023-01949-9.

Murray, T., Binetti, N., Venkataramaiyer, R., et al., 2024. Expression perceptive fields explain individual differences in the recognition of facial emotions. Commun. Psychol. 2, 62. https://doi.org/10.1038/s44271-024-00111-7.

Pinkham, A.E., Sasson, N.J., Calkins, M.E., et al., 2008. The other-race effect in face processing among African American and Caucasian individuals with schizophrenia. Am. J. Psychiatry 165 (5), 639–645. https://doi.org/10.1176/appi.ajp.2007.07101604.

Roubtsova, N., Parsons, M., Binetti, N., Mareschal, I., Viding, E., Cosker, D., 2021. EmoGen: quantifiable emotion generation and analysis for experimental psychology. arXiv [preprint], pp. 1–21. https://doi.org/10.48550/arXiv.2107.00480.

Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York.

Stolk, A., Verhagen, L., Toni, I., 2016. Conceptual alignment: how brains achieve mutual understanding. Trends Cogn. Sci. 20 (3), 180–191. https://doi.org/10.1016/j.tics.2015.11.007.