Neural correlates reveal separate stages of spontaneous face perception

Check for updates
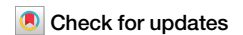
Amanda K. Robinson [1,2,3] ✉, Greta Stuart[1], Sophia M. Shatek[3,4], Adrian Herbert [1] & Jessica Taubert [1]

The human brain rapidly detects faces, even in inanimate objects—a phenomenon known as face pareidolia. While this illusion reveals the automaticity of face detection, it also presents a paradox: how does the brain process stimuli that are simultaneously faces and objects? Here, we combined behavioral experiments with electroencephalography to track the temporal dynamics of face pareidolia processing. Using a large stimulus set of human faces, objects containing illusory faces, and matched control objects, we show that perception of face pareidolia is remarkably flexible and task dependent. When making spontaneous similarity judgments, participants ($N = 338$) perceived illusory faces as intermediate between faces and objects. However, in explicit categorization tasks ($N = 22$), the same stimuli were predominantly classified as objects, while rating face-likeness ($N = 20$) bolstered the representation of face-like features. Neural responses ($N = 20$) tracked this perceptual flexibility: early visual processing (90–130 ms) correlated with face-like judgments, while later activity (150–210 ms) aligned with object categorization. This temporal progression demonstrates how the brain maintains multiple levels of representation, integrating early face detection with subsequent object recognition to support flexible behavior. Our findings demonstrate that face pareidolia exemplifies the brain's capacity to resolve perceptual ambiguity through dynamic processing, with task demands determining how competing representations contribute to perception.

The primate brain responds differently to faces and objects. This ubiquitous finding, which transcends all recording[1–3], imaging[4–7], and causal[8–10] methods, has led to an unresolved debate: do faces engage distinct neural mechanisms, separate from those processing objects[11]? Or, alternatively, do visually evoked responses reflect a distributed code that can universally classify any visual stimulus as a function of image properties[12]? These questions, while central to our understanding of primate vision and brain topography, have been largely rendered intractable, in part because it is difficult to decouple a face from the image properties that typically define a face. Difficult but not impossible.

Face pareidolia is the common experience of perceiving a face in an otherwise inanimate object[13–15]. Interestingly, while these stimuli are easily perceived as face-like, their image properties are more typical of inanimate objects[16–18]. Thus, examples of face pareidolia provide a rare opportunity to understand how image properties contribute to the neural representation of faces and objects. For instance, examples of face pareidolia have been used to show that the brain's response to a visual stimulus evolves over time[19–21]; patterns of brain activity first encode examples of face pareidolia as being more similar to faces, and then more similar to objects[21]. This is consistent

with a system that has distinct mechanisms underscoring distinct, potentially parallel, functions. To date, however, there is no evidence that the evolving neural signature of face pareidolia has behavioral consequences. In other words, when we experience the face pareidolia illusion, do we first see a face and then an object, or do we perceive a mixture of both in a stimulus-dependent manner?

Studies of face pareidolia at the behavioral level have been narrowly focused on the perception of illusory faces, often asking participants to rate how face-like an image is on an ordinal scale[21–23] or to locate face-like patterns in natural scenes[15,24] or pure noise[25,26]. These tasks motivate participants to actively search for evidence of face-like features, potentially biasing behavior. Further, these responses correlate with brain activity late in the time course (see Romagnano et al.[27]; Wardle et al.[21]), suggesting that they index cognitive decisions rather than early sensory processes or spontaneous behaviors that occur without awareness. What is needed are better behavioral markers of the face pareidolia illusion, where participants are not instructed to search for, or evaluate, facial attributes.

To address this knowledge gap, we used an odd-one-out triplet task to measure perceived dissimilarity among a large number of images comprised

[1]School of Psychology, The University of Queensland, Brisbane, Australia. [2]Queensland Brain Institute, The University of Queensland, Brisbane, Australia. [3]School of Psychology, University of Sydney, Sydney, Australia. [4]Department of Experimental Psychology, University of Oxford, Oxford, UK. ✉e-mail: amanda.robinson@uqconnect.edu.au

of 100 human faces, 100 objects with illusory facial features (hereafter referred to as illusory faces) and 100 matched objects (Fig. 1A). This task has been used to characterize the latent featural dimensions underlying visual recognition, without constraining or guiding participant responses[28,29]. Assuming the face pareidolia illusion is perceived spontaneously, even when participants are not instructed to look for faces, we predicted that illusory faces would be perceived as more similar to each other than to their non-face objects counterparts, despite being matched for semantic content. Additionally, for all 300 stimuli, we collected face-like ratings, predicting that the ratings given to illusory faces in this behavioral context would better reflect their illusory face identity (Fig. 1B), and responses during a face-object categorization task, predicting that responses to illusory faces in this behavioral context would better reflect their veridical object identity (Fig. 1C). Then we collected time-resolved visual evoked responses to the 300 stimuli using EEG (Fig. 1D). Our goal was to properly contextualize human behavior towards examples of face pareidolia, relative to both real faces and ordinary objects, and then leverage those observations to better understand how the evolving neural representation of face pareidolia supports behavioral responses.

Our primary goal was to determine whether behavioral responses capturing the illusory face identity in examples of face pareidolia would correlate with brain activity at an earlier time point than behavioral responses that only capture the veridical object identity in examples of face pareidolia[21,30]. We also hypothesized that behavioral responses in the odd-one-out triplet task would distinguish between nested stimulus pairs (i.e., an illusory face and a matched object with no face) at earlier stages of processing than the speeded categorization task. Collectively, these observations would suggest that the mechanisms responsible for detecting faces, i.e., distinguishing real and illusory faces from other kinds of visual stimuli—are distinct from those responsible for recognizing objects.

Methods

Code, stimuli and behavioral data are available at https://doi.org/10.5281/zenodo.15833508. EEG data are available at https://doi.org/10.18112/openneuro.ds005642.v1.0.0. This study was not preregistered.

Stimuli

Stimulus images consisted of human faces, illusory faces and matched non-face objects. There were 300 stimuli: 100 exemplars each of human faces, illusory faces and matched objects. The same stimuli were used in all three behavioral tasks and the EEG experiment.

Spontaneous dissimilarity task

In this task, participants rated the similarity between the 300 experimental stimuli using a triplet odd-one-out task[28,29]. Participants were 338 undergraduate students who participated in return for course credit (self-reported gender: 221 females, 107 males, 4 non-binary, 6 prefer not to say; median age 19 years, range 17–48 years). An additional 10 participants started but did not complete the experiment (<40% trials completed), and their data were not analyzed. This study was approved by the University of Queensland ethics committee (number 2021/HE002275) and informed consent was obtained from all participants. The experiments were programmed in jsPsych[31] and hosted on Pavlovia[32]. On each trial, three experimental stimuli were presented simultaneously, and participants were asked to choose the odd one out by clicking on the stimulus (Fig. 1A). Stimuli were presented equidistant from fixation in a triangle pattern. There were 300 trials in the experiment.

There was one main round of data collection with two subsequent rounds to ensure enough data coverage across all stimuli. First, we collected judgements from $N = 328$ participants using all 300 stimuli, with stimulus combinations chosen randomly on each trial. After collating the results, we assessed which pairs of stimuli had never been presented together (67 pairs) and collected 5 more participants by ensuring these stimulus pairs were included and excluding pairs with the highest number of presentations (68 pairs). In a final round, we assessed which stimulus pairs had only appeared

once (372 pairs); we collected 5 more participants, including these pairs and excluding pairs presented the most (397 pairs with more than 13 presentations). In total, there were 101,397 trials, with each of the 404,550 stimulus pairs presented at least once.

All trials from all participants were collated, and the behavioral responses were used to construct a representational dissimilarity matrix (RDM). For each trial, dissimilarity was calculated for the pairs of stimuli (3 separate pairs for the 3 distinct stimuli). The stimulus chosen as the odd-one-out was coded as dissimilar from each of the other two stimuli (values of 1), and the two other stimuli were coded as similar (value of 0). The dissimilarity of each stimulus pair (e.g., face #1 vs illusory face #17) was calculated as the mean dissimilarity value for all trials in which those two stimuli were presented together. These mean values were used to construct a 300 ×300 RDM (Fig. 2A, left).

Face-like task

Behavioral ratings were collected for the 300 stimuli. Participants were 20 undergraduates from the University of Queensland (self-reported demographics: 18 females, 2 males; median age 18.5 years, range 17–42 years) who participated in return for course credit. This study was approved by the University of Queensland ethics committee (number 2021/HE002275), and informed consent was obtained from all participants. Participants were shown each stimulus in turn, in random order, and asked to "Rate how easily you can see a face in this image" on a scale of 0–10 (Fig. 1B), as in previous work[21–23]. The experiment was programmed in Qualtrics. Data from all participants were collated, and the group mean face-like score was calculated for each stimulus. A 300 × 300 RDM was constructed using Euclidean distance of face-like scores for each pair of images (Fig. 2A, middle).

Face-object categorization task

In the final behavioral task, we used a forced-choice categorization task. Participants ($N = 22$; self-reported demographics: 16 females, 6 males; median age 22 years, range 19–30 years) were recruited from the University of Queensland in return for payment in the form of gift cards (AUD$20). An additional participant completed the study but was not included in the analyses due to equipment failure. This study was approved by the University of Queensland ethics committee (number 2021/HE002275) and informed consent was obtained from all participants. They completed an experimental session in the laboratory.

On each trial, a fixation cross was presented for 300 ms, followed by a stimulus image, and finally, a Mondrian mask image was presented until the response (Fig. 1C). Participants were asked to press a button to indicate if the stimulus was a face or an object. Stimuli were presented for 33.33 or 100 ms, designed to tap into different stages of processing. There were 1200 trials in total, with two repeats of each of the 300 stimuli per stimulus duration, presented in random order. Participants were given a break every 300 trials. The experiment was programmed in Psychopy[32] on a 1920 × 1080 VPixx monitor set at a refresh rate of 60 Hz. Stimuli and masks were presented at a size of 8 ×8 degrees of visual angle.

For each participant, the proportion of "face" categorization responses was calculated for each stimulus and presentation condition. Data from all participants were then collated, and the group mean face-like score was calculated for each stimulus. There was no difference in the responses between the two stimulus duration conditions ($BF_{10} = 0.006$, $t_{21} = 0.23$, $p = 0.817$, $d = 0.05$, 95% $CI_{diff} = [−0.01\ 0.01]$) so we took the mean of both durations. A 300 × 300 RDM was then constructed using Euclidean distance of face categorization scores for each pair of images (Fig. 2A, right).

EEG experiment

Participants viewed stimuli that appeared centrally at fixation while their electroencephalography was used to measure neural activity from the scalp. Participants were 20 adults recruited from the University of Queensland (self-reported demographics: 16 females, 4 males; median age 22.5 years, range 18–30 years) and were compensated for their time at a rate of AUD $20 per hour. An additional participant completed the study but was not
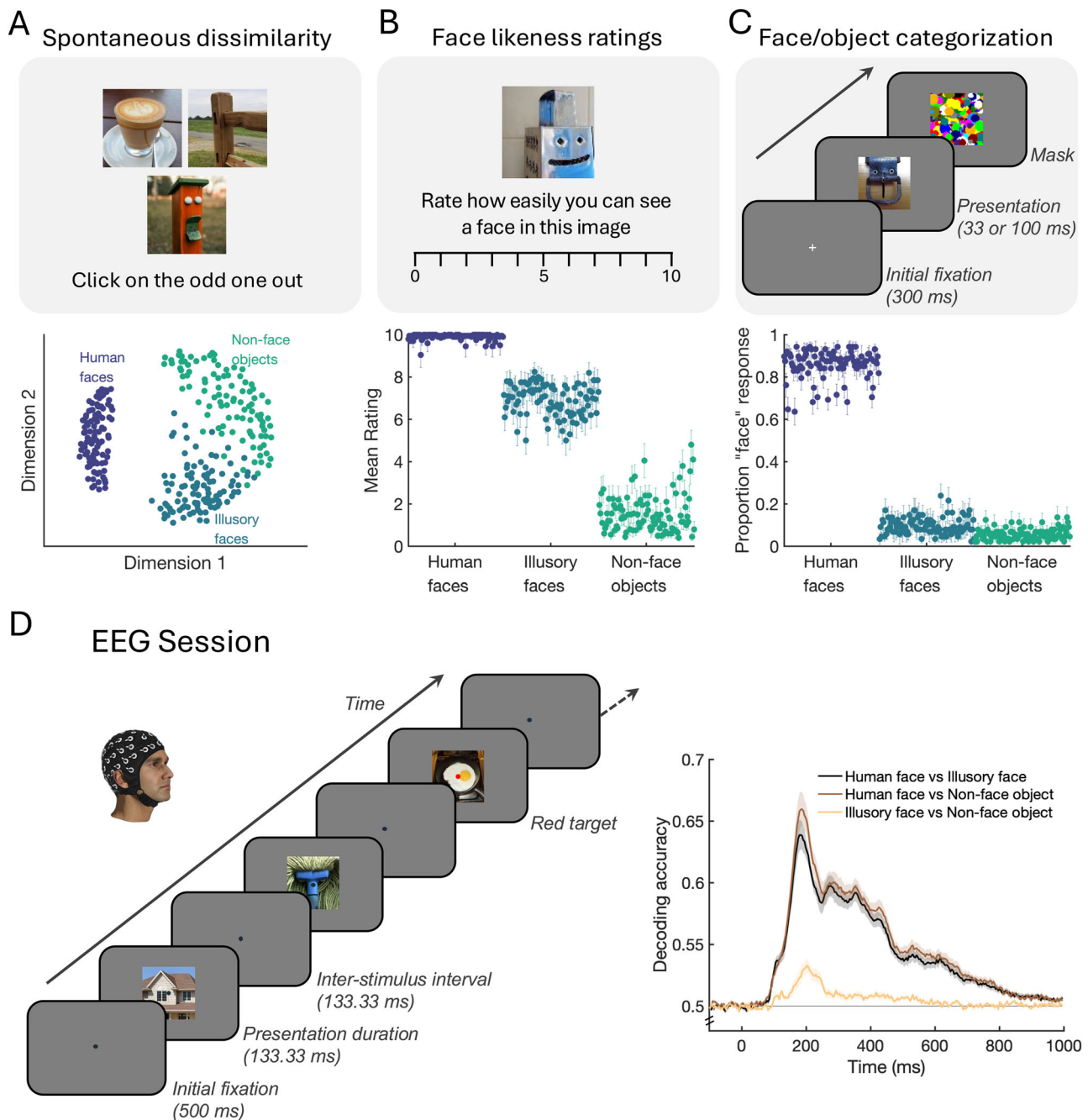
Fig. 1 | Experimental design. A Triplet odd-one-out task, capturing spontaneous dissimilarity judgements across images in the stimulus set. Multi-dimensional scaling of the dissimilarity scores revealed three clusters for the categories of human faces, illusory faces and objects. **B** Face-like ratings task. Participants were asked to give a face-like rating to each stimulus on a scale from 0 to 10. Illusory faces were rated between objects and human faces. **C** Categorization task. In each trial, participants were presented with a stimulus that was backward masked and asked to categorize it as a face or object. Results revealed illusory faces were largely judged as objects. **D** Example sequence timeline from EEG experiment (left) and results from decoding the three categories (right). Participants viewed sequences of stimuli at 3.75 Hz while their neural responses were measured with EEG. Throughout the session, they performed an orthogonal fixation color change detection task. Mean decoding accuracy is shown for category pairs over time, compared to chance (50%), showing a similar pattern of results to previous work[22].

included in analyses due to equipment failure. This study was approved by the University of Queensland ethics committee (number 2020/HE003101), and informed consent was obtained from all participants. All participants reported normal or corrected-to-normal vision.

Stimuli were presented using Psychopy[32] at ~4 × 4 degrees of visual angle on an LCD monitor (VIEWPixx 3D, VPixx Technologies; 1920 × 1080 pixels, 22.5-inch, 120 Hz refresh rate). Images were presented in sequences of 150 stimuli such that two adjacent sequences contained each of the 300 experimental stimuli once, in random order. Every sequence

began with a fixation dot for 500 ms, then stimuli were presented one after another for 133.33 ms with an inter-stimulus interval of 133.33 ms (i.e., image presentation at a rate of 3.75 Hz). Across the experiment, there were 70 sequences totaling 10,500 trials, consisting of 35 repeats for each of the 300 stimuli.

During the experimental session, participants were asked to maintain fixation on the dot that appeared in the center of the screen in black, detect when it turned red (Fig. 1D) and indicate detection by button press. This task was designed to be orthogonal and irrelevant to the stimuli. The rapid
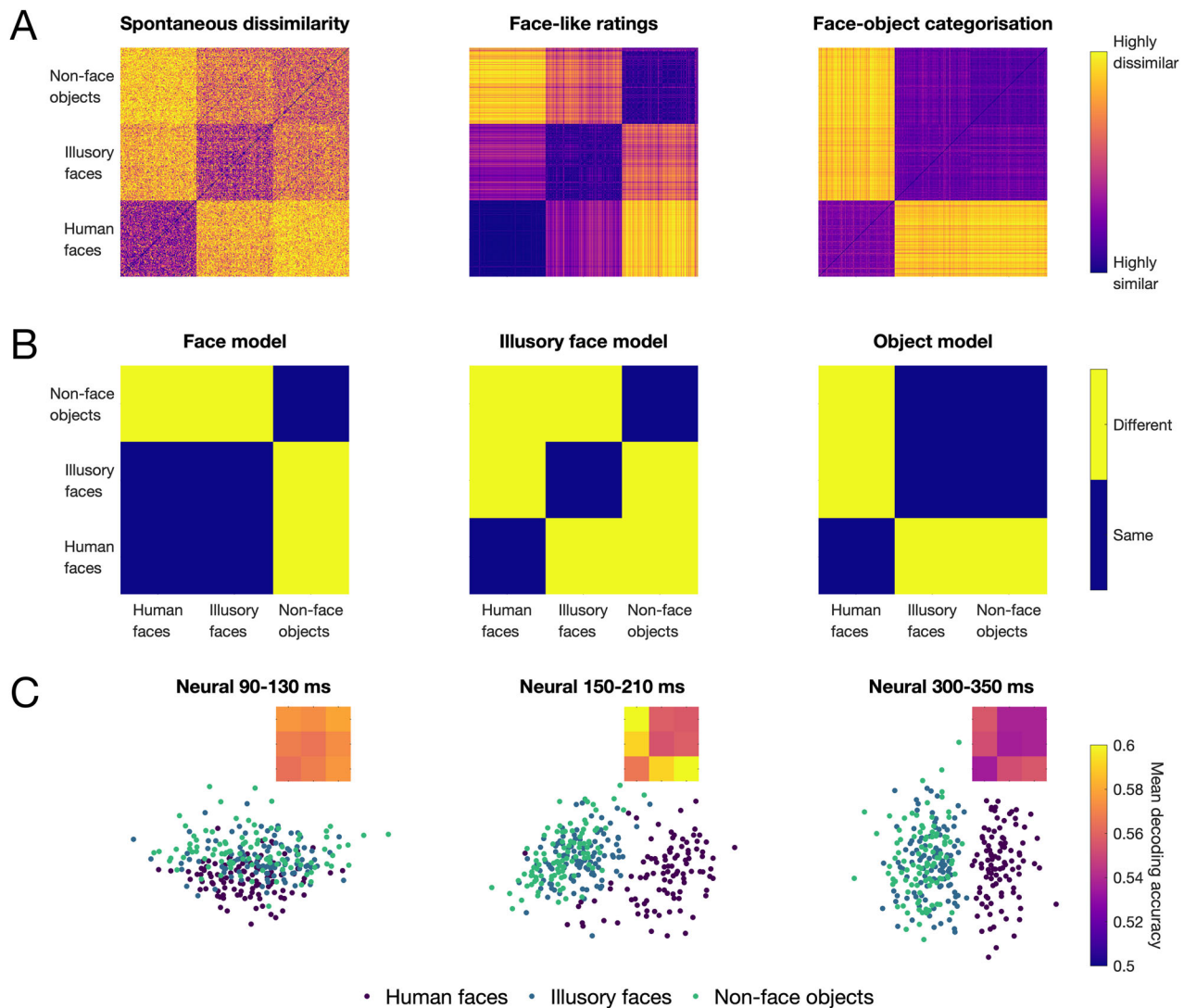
Fig. 2 | Representations of human faces, illusory faces and objects in behavior, categories and neural responses. A Representational dissimilarity matrices (RDMs) based on behavior from the spontaneous dissimilarity ($N = 338$), face-like ($N = 20$) and categorization tasks ($N = 22$). **B** Face-object category models that vary according to the category assigned by illusory faces: illusory faces are coded equivalent to human faces (face model), coded as a separate third category (illusory face model), and coded equivalent to objects (object model). **C** Neural representations of the 300 experimental stimuli from three different stages of processing using multi-dimensional scaling. Inset plots show mean neural decoding RDMs from that time period, downsampled to the category level.

stimulus presentation, random image sequences and orthogonal task reduced the likelihood of participants moving their eyes in a stimulus-specific manner.

EEG recording and preprocessing

EEG data were continuously recorded from a 64-electrode BioSemi system, arranged in the international 10–20 system for electrode placement[33], digitized at a sample rate of 1024 Hz. The EEGLAB toolbox[34] was used to preprocess the data offline. First, we re-referenced to channel Cz, then filtered the data using a Hamming windowed sinc FIR filter with a high pass of 0.1 Hz and a low pass of 100 Hz, as in our previous work[35,36]. Following these steps, noisy electrodes were identified using joint probability and were reconstructed using spherical interpolation if they exceeded 5 standard deviations from the average (mean number interpolated = 0.25, min = 1, max = 3). A common average reference was then applied, and data were downsampled to 256 Hz. Finally, epochs were created for each stimulus presentation from [−100 to 1000 ms] relative to stimulus onset, and baseline corrected. No other pre-processing or data cleaning was performed.

Neural decoding

To investigate how the perception of face pareidolia unfolds over time, we assessed the neural representations of human face, illusory face and matched non-face object stimuli. Multivariate pattern analysis, or neural decoding, was applied to the time-resolved EEG data to discriminate how different stimuli evoked different patterns of neural activity over the scalp[37,38]. For each time point (3.90 ms time resolution) and participant, we assessed stimulus-specific representations by training a classifier to discriminate between neural activity associated with two experimental stimuli and testing on held-out data for the same stimuli. Decoding was implemented using the CoSMoMVPA toolbox[39]. Data were pooled across the 64 EEG sensors, and we tested the ability of a linear discriminant analysis (LDA) classifier to discriminate between the patterns of neural responses associated with each stimulus. A 35-fold cross-validation procedure was used, with each fold containing 2 independent trial sequences (one repeat of each stimulus). All pairs of combinations for the 300 stimuli (e.g., humanface1 vs illusoryface2, illusoryface87 vs matchedobject4) were decoded, resulting in 44,850 unique contrasts across time per participant. Classifier accuracy was calculated as the mean proportion of correct classifier predictions across all folds.

Above-chance group mean decoding accuracy (above 50%) was considered evidence of stimulus information in the neural signals.

Representational similarity analyses

To investigate the relationship in the structure of stimulus representations between the neural responses measured via EEG and perception measures via behavioral tasks, we used representational similarity analyses (RSA)[12]. RSA involves a neural-behavior comparison that is abstracted away from task-specific or methodology-specific responses, instead focusing on the relationships between stimulus representations. This set of analyses allowed us to assess the content of information within neural representations that relates to behavior.

Using the neural and behavioral results, we constructed representational dissimilarity matrices (RDMs), which quantified the similarity between each stimulus. Each of these RDM models was a 300×300 matrix of dissimilarity for each of the 300 stimuli with each other stimulus, using the relevant neural or behavioral measure. The RDMs were symmetrical across the diagonal, with 44,850 unique values.

For each behavioral task, a dissimilarity matrix was constructed based on the difference in mean behavioral responses across stimulus pairs (Fig. 2A). As a comparison, we constructed three face-object category models: a face model that codes illusory faces as faces, an object model that codes illusory faces as objects, and an illusory face model that codes illusory faces as a separate third category from faces and objects (Fig. 2B). All behavioral and category models were significantly correlated ($rho > 0.097$, ps < 0.001; see Supplementary Fig. 1 and Supplementary Table 1 for relationships between models)).

Neural RDMs used decoding accuracy for each pair of stimuli at each time point (3.90 ms temporal resolution). Separate 300×300 neural RDMs were constructed for each time point and participant, where each cell contained the mean decoding accuracy between two stimuli. The behavioral RDMs were based on the group mean dissimilarity scores from the three behavioral experiments. We also constructed three additional stimulus models based on the stimulus category: the face model, which classed illusory faces and faces as distinct from objects; the object model, which classed illusory faces and objects as distinct from human faces; and the illusory face model, which classed illusory faces as distinct from both human faces and objects.

Using RSA, we investigated how neural representations related to behavioral judgements. Neural RDMs per participant were correlated with each behavioral RDM using Spearman correlation to assess similarity of the lower diagonals of the RDMs (i.e., the unique pairwise values), for every time point. This allowed us to assess how neural information might inform overall perception. Correlations were performed for each EEG participant separately, and the mean was calculated across the group.

To establish a boundary for model performance, we calculated the noise ceiling, an estimate of how well any model could explain the neural data, considering its noise and variability. At each time point, we rank-transformed each participant's neural RDM and Spearman correlated with the mean rank-transformed RDMs from all other participants. The mean correlation across participants provided the lower bound estimate of the noise ceiling[40].

Statistical testing

To assess neural-behavior correlations, we used Bayesian statistics to determine the evidence for the alternative relative to the null hypotheses[41–45]. For RSA analyses, the alternative hypotheses of above- and below-zero correlations were tested. Data distribution was assumed to be normal, but this was not formally tested. We used the 'BayesFactor' package in R[46]. Bayes Factors were calculated using a JZS prior, centered around chance correlations of zero[44] with a default scale factor of 0.707, meaning that for the alternative hypotheses of above- and below-zero correlations, we expected to see 50% of parameter values falling within -0.707 and 0.707 standard deviations from chance[43,44,47,48]. A null interval was specified as a range of effect sizes between -0.5 and 0.5[49].

A Bayes Factor (BF_{10}) is the probability of the data under the alternative hypothesis relative to the null hypothesis. We consider $BF_{10} > 3$ as evidence for the alternative hypothesis (above-chance decoding and reliable correlations). We interpret $BF_{10} < 1/3$ as evidence in favor of the null hypothesis[43,50].

Reporting summary

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

Results

### Behavioral tasks reveal context-dependent processing of illusory faces

Our investigation into the neural processing of illusory faces for visual recognition employed three complementary behavioral tasks, each probing different aspects of behavioral recognition: (1) perceptual similarity, (2) face-like appearance, and (3) face-object categorization. To this end, separate groups of participants performed tasks of spontaneous dissimilarity judgments ($N = 338$), explicit face-likeness ratings ($N = 20$), and speeded face-object categorization ($N = 23$) based on the same 300 stimuli. These results unveiled a remarkable flexibility in how the human visual system builds multiplexed representations of illusory faces.

First, in the spontaneous dissimilarity task, participants performed odd-one-out judgments among triplets of stimuli (Fig. 1A). As predicted, the resulting multidimensional similarity space revealed a striking organization: while human faces and non-face objects formed distinct clusters, illusory faces fell midway between them (see Supplementary Fig. 2 for analyses showing higher between than within-category dissimilarity). Although the illusory faces were positioned closer to the object cluster than the face cluster, there was little overlap between the illusory faces and matched objects. This organization suggests that illusory faces have an inherent dual nature[18,51] and are spontaneously perceived as distinct from ordinary objects even when participants are not prompted to look for face-like features.

Next, the face-likeness rating task captured a different aspect of the graded nature of illusory face perception. Importantly, human faces garnered near-ceiling face-like ratings ($M = 9.90$, SE = 0.04), while non-face objects received minimal face-like scores ($M = 1.54$, SE = 0.27), indicating that participants understood the task and aligning these results with previous studies[21–23]. Interestingly, illusory faces elicited robust face-like ratings ($M = 6.93$, SE = 0.45) that were significantly lower than human faces ($BF_{10} = 1.68 \times 10^4$, $t_{19} = -7.03$, $p < 0.001$, d = 1.57, 95% $CI_{diff} = [-3.86 -2.09]$) but also significantly higher than objects ($BF_{10} = 1.60 \times 108$, $t_{19} = 13.14$, p $< .001$, d = 2.94, 95% $CI_{diff} = [4.52 6.24]$). The observation that the mean face-like ratings for illusory faces were above 5 ($BF_{10} = 84.84$, $t_{19} = 4.30$, $p < 0.001$, $d = 0.96$, 95% $CI_{illusory} = [5.99 7.86]$) reflects a bias towards rating illusory faces as being more face-like than object-like. This bias is consistent with previous studies using this approach[21–23] and confirms the perception of facial features in this large set of illusory face stimuli.

Finally, the speeded categorization task revealed a different behavioral response profile. When forced to make binary face-object decisions, participants predominantly classified illusory faces as objects rather than faces (proportion face response $M_{illusoryface} = 0.10$). Human faces were classified more frequently as faces than illusory faces ($M_{humanface} = 0.86$; $BF_{10} = 2.91 \times 1012$, $t_{21} = 20.77$, $p < 0.001$, $d = 4.43$, 95% $CI_{diff} = [0.69 0.84]$). However, there was no evidence for a difference in rate of face responses for illusory faces and non-face objects ($M_{nonfaceobject} = 0.06$; $BF_{10} = 1.04$, $t_{21} = 1.91$, $p = 0.070$, $d = 0.41$, 95% $CI_{diff} = [-0.004 0.08]$). These findings suggest that when instructed to make categorical decisions, the veridical object identity takes precedence over illusory facial characteristics and illusory faces are reported as being mere objects. In sum, by employing the same stimuli over three independent behavioral tasks, each time collecting a new sample of participants, we show that the behavioral context changes how participants respond to face pareidolia. On the one hand, certain tasks will capture and even augment the face-like appearance of the illusory identity, while on the

Fig. 3 | Neural representations of illusory faces and their behavioral relevance. Time-varying correlations between neural responses (*N* = 20), behavior and visual category models. **A** Responses on three different behavioral tasks reflected different dynamics in the neural signal. **B** Category models that considered illusory faces either equivalent to human faces, equivalent to objects, or a distinct third category correlated differentially with neural representations over time. Specifically, the face model had the highest correlation for the first stage of processing, but was rapidly overtaken by the object model. These results indicate a change in illusory face processing over time, where illusory faces initially and briefly resemble human faces but subsequently resemble objects. Noise ceiling reflects the lower bound of the expected RDM correlation based on the noise in the data. Error bars (shaded areas) reflect one standard error of the mean.
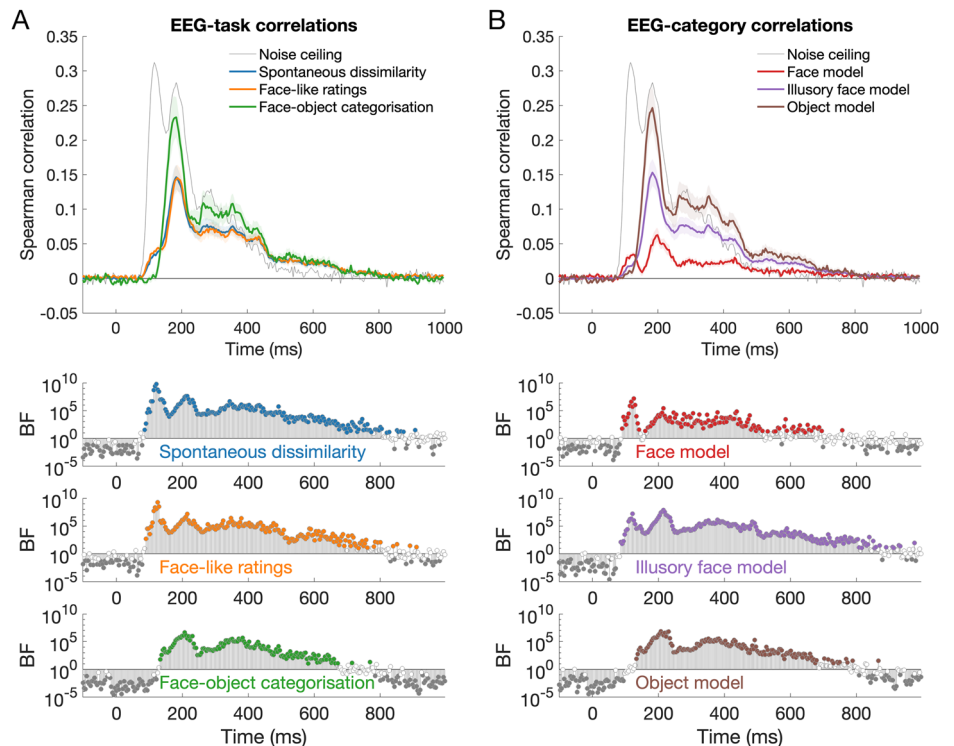
other hand, other tasks will unbind and ignore the illusory identity in favor of the true, veridical identity. Importantly, when we measured perceptual similarity using the triplet odd-one-out task, we confirmed that when unprompted by instruction and unconstrained by time, participants spontaneously perceived illusory faces in examples of face pareidolia and reported them as being distinct from matched non-face objects.

Neural dynamics track behavioral flexibility

Having established that the perceived identity of illusory faces is malleable, our key question is whether these different identities are represented at different stages of neural processing. To bridge neural processing with behavior, we employed representational similarity analysis (RSA) to map the dynamics of stimulus processing. Representational dissimilarity matrices (RDMs) were constructed for each behavioral task and three category models (Fig. 2A/B; see Supplementary Fig. 1 and Supplementary Table 1 for relationships between models). Additionally, RDMs were constructed for the neural representations at each time point (representative time windows shown in Fig. 2C). Neural RDMs at each time point were then correlated with each task and category RDM. This approach revealed rich temporal patterns linking neural representations to behavioral judgments across our tasks (Fig. 3A).

The earliest neural processing stage (90–130 ms) showed reliable correlations with spontaneous dissimilarity judgments and face-likeness ratings, but notably not with categorical face-object decisions (Fig. 3A). This early window appears to capture initial face-like processing of illusory stimuli, as confirmed by stronger correlations with a face-based category model compared to an object-based model (Fig. 3B). However, a dramatic shift occurred in the 150–210 ms window; neural patterns showed the strongest correlation with face-object categorization behavior (Fig. 3A) and the object-based category model (Fig. 3B). This temporal transition indicates that a rapid re-coding of illusory faces takes place, shifting from an initial face-like representation to strict object representation. In the third distinct time window (300–350 ms), the neural information patterns again favored face-object categorization over spontaneous dissimilarity judgements and face-like scores, but with lower fidelity than during the second time window. Complementary variance partitioning analyses revealed

similar changes over time, while also confirming that each of the three behavioral tasks accounted for unique and common variance in the neural signal (see Supplementary Figs. 3–5). This persistence of object-like categorical processing, rather than a return to perceptual similarity, suggests that the brain maintains and refines canonical category representations even in late processing stages. Overall, this evolving neural signature helps explain the flexibility in human behavior in regard to face pareidolia; when we see an example of face pareidolia, the brain is equipped to build and maintain a multiplexed representation of that stimulus.

Faces in objects: behavior reflects multidimensional brain responses

Beyond larger categorical distinctions, we refined our analysis to examine the 100 pairs of nested stimuli included in the design. Each pair was comprised of an illusory face and a matched non-face object (e.g., a cookie with an illusory face and a cookie without an illusory face). Subsetted RDMs showing only the illusory face and non-face object comparisons are presented in Fig. 4A, with the matched pairs on the diagonal. Figure 4B shows the neural-behavior correlations for the nested pairs across the three tasks. The neural dissimilarity between stimulus pairs correlated with spontaneous dissimilarity judgments during early processing (105–136 ms, median $BF_{10} = 11617$), whereas the face-likeness ratings showed no reliable correlation, and the face-object categorization task exhibited a neural-behavioral relationship that was later in the time course (156–207 ms, median $BF_{10} = 10.55$). This temporal dissociation provides further evidence that tasks yield different insights into image separability in the brain. Notably, the spontaneous triplet task proved particularly valuable by capturing multiple levels of image dissimilarity (i.e., category-level and exemplar-level) without explicit instruction, demonstrating its effectiveness as a measure of natural visual processing.

Discussion
Behavioral flexibility and task dependencies
In this study, we explored the neural representation of illusory faces and their relationship to human behavior across a range of tasks. A key advance of this project was the implementation of an unbiased task to measure
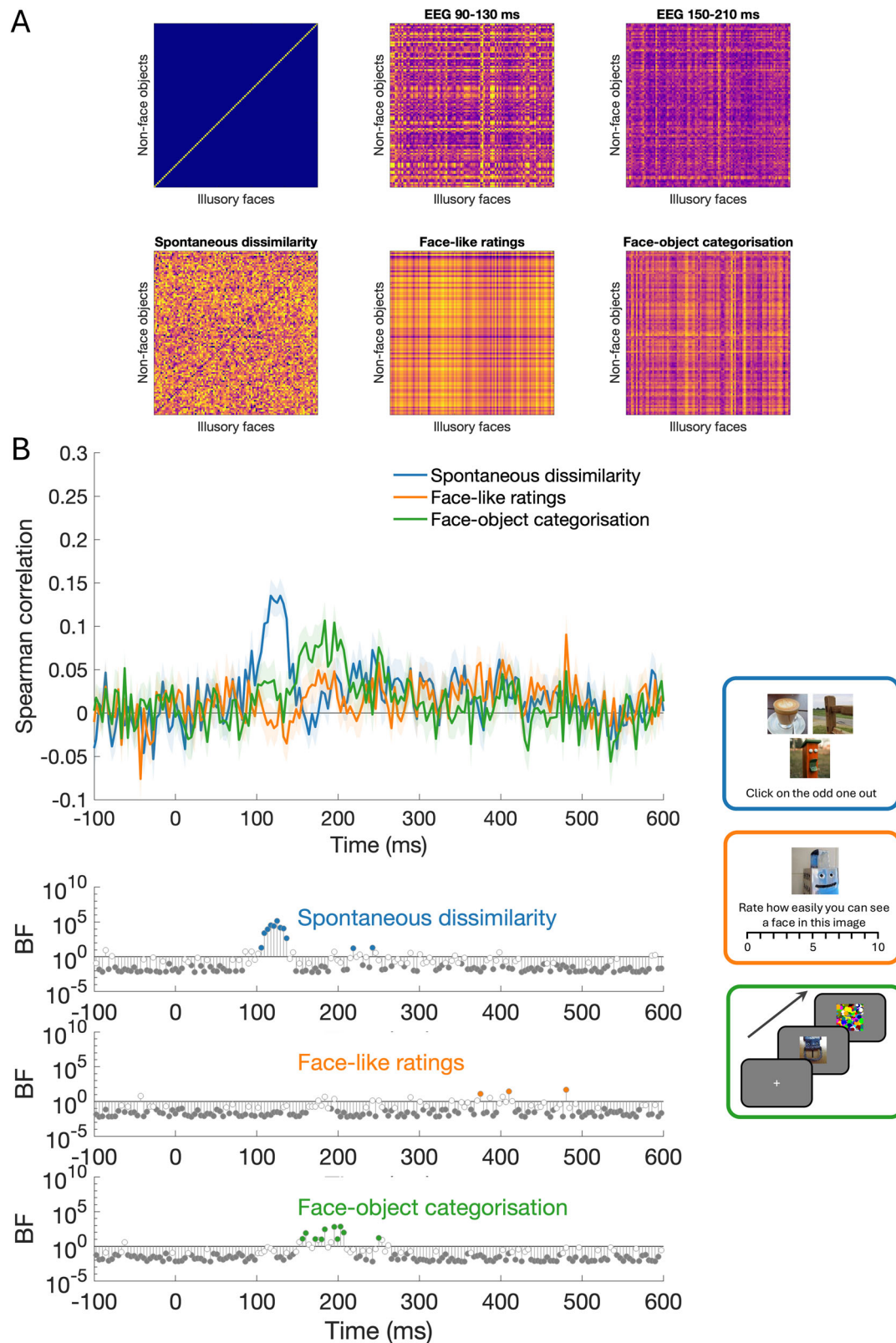
Fig. 4 | Neural representations of illusory faces and their behavioral relevance. A RDM subset showing only the illusory face x non-face object comparisons in the neural responses (*N* = 20) and for each behavioral task. Top left RDM shows the exact cells used for the matched pair analysis in yellow. Note that the RDMs are not symmetrical because illusory faces and non-face objects are represented on different axes. The diagonal for matched pairs is clear in the spontaneous dissimilarity task, but not in the other tasks. **B** Time-varying correlations between neural responses and behavioral tasks for the 100 nested stimulus pairs, consisting of illusory faces and matching non-face objects. Neural information was correlated with responses on the spontaneous dissimilarity task (blue) from early stages of processing, and the face-object categorization task (green) later in the time course. No consistent, reliable correlation was captured for the face-like ratings (orange). Error bars (shaded areas) reflect one standard error of the mean.

perceptual dissimilarity among faces, illusory faces, and matched objects within a multidimensional space. This approach allowed us to confirm that human participants spontaneously perceive illusory faces in ambient images of objects that coincidentally resemble faces. To gain a comprehensive understanding of the behavioral dynamics, in separate experiments, we asked participants to: (1) rate the face-like appearance of each stimulus and (2) rapidly categorize each stimulus as either a face or an object. As expected, these conventional behavioral tasks yielded contrasting biases: participants rated illusory faces as more face-like than object-like in the ratings task, whereas participants categorized illusory faces as more object-like than face-like in the categorization task. These findings underscore the considerable flexibility in the perception of illusory faces at the behavioral level, which one might expect given their ambiguous and illusory nature. The next step was to connect these behavioral patterns to time-resolved neural activity using multivariate analysis methods.

Temporal dynamics of illusory face perception

Using the same large set of images (i.e., 300 face and non-face stimuli), we employed EEG to first replicate the finding that the neural representation of face pareidolia shifts from face-like to object-like over time[19–21]. Then, by converging behavioral and neural evidence, we discovered that the behavioral markers that captured the illusory face identity in examples of face pareidolia (i.e., behavior in the triplet task and the ratings task) correlated with the earliest evoked responses. In contrast, behavior in the face-object categorization task, which emphasized the veridical object identity in examples of face pareidolia, correlated with later evoked responses. This delay is somewhat paradoxical because the face-object categorization task was the only speeded task. Rather than simply reflecting response speed, this temporal pattern suggests that perception of face pareidolia unfolds through distinct computational stages, with early processes supporting face detection and later stages mediating object recognition. One possibility is that the face-selective cortex is more excitable than object-selective regions. This would lead to the rapid propagation of a neural signal reflecting the illusory face identity before knowledge about the object identity could be extracted. Interestingly, while the correlations occur at different onsets, it is clear that the associations between behavior and the neural time course were maintained over extended and overlapping periods of time. Notably, early face-identity signals persist even after object-identity processes dominate, suggesting the brain preserves initial interpretations rather than over-writing them. This provides critical insight into the brain's capacity to build and maintain multiple independent representations in parallel.

The neural time-course underlying the spontaneous perception of face pareidolia

Crucially, the results of the untargeted triplet odd-one-out task demonstrate that the facial features in illusory objects were spontaneously perceived and used to make similarity judgements. While the neural-behavior correlations that incorporate all 300 stimuli (Fig. 3) were similar for the triplet and face ratings task, likely reflecting broad category-level differences, Fig. 4's exemplar-specific analysis shows that the triplet task uniquely captured information about semantically-matched object pairs. This information was not captured by the explicit ratings or categorization behaviors. This demonstrates that spontaneous dissimilarity judgements yield higher dimensionality than explicit ratings or categorization behaviors with an enriched feature-specific signal capturing the similarity between object exemplars from the same semantic categories ("nested object pairs"). The size of the stimulus set employed in our study was important for detecting these subtle differences, enabling a nuanced understanding of how targeted or untargeted task context influences the perception of ambiguous stimuli. This task-dependent perceptual flexibility supports our previous work[18,51] showing that behavior towards images containing two identities (i.e., objects with illusory faces) will differ depending on the task at hand (e.g., object-detection or face-detection).

Behavioral relevance as an organizing principle in the visual cortex

Our findings align with theoretical frameworks that argue the organization of the visual cortex reflects behavioral goals and stimulus affordances, not putative stimulus categories[52]. Our findings indicate that a single stimulus may evoke multiple neural representations that are maintained over time. This capacity for building multiplexed representations would afford maximum behavioral flexibility. For example, when we encounter face pareidolia, information about facial features is extracted and becomes available for any task that makes the face relevant. But, shortly thereafter, information about the object identity is computed for any task that makes the object relevant. It follows that, because examples of face pareidolia have two distinct identities (i.e., a face and an object), possibly represented by distinct mechanisms[1,53,54], face pareidolia is a particularly useful tool for probing multiplexed representations. That said, multiplexed representations would not necessarily be exclusive to face pareidolia because many, if not all, visual objects could have latent identities, depending on a person's experience and task demands. For example, when we see a banana, our brain might extract the visual features that make it look like a fruit and, thus, build a representation similar to other fruit,s but our brain might also extract the visual features that make a banana look 'tossable' and, thus, build a representation similar to other potential projectiles. This perspective emphasizes the need to consider behavioral goals and stimulus affordances, but also their flexibility, in understanding how the brain represents visual stimuli.

Limitations

This study advances our understanding of illusory face perception, yet important questions remain about the underlying mechanisms. Future research could investigate how different tasks causally influence neural processing to test the malleability of illusory face representations with top-down processing. Such work would provide an important contrast to the bottom-up aspects of visual processing we tested here and could reveal whether the fundamental representational organization we observe can be dynamically reconfigured by task goals. Additionally, it would be interesting to investigate individual differences in the propensity to perceive illusory faces in specific exemplars[23]. Understanding these variations could provide deeper insights into how neural responses shape perception and might reveal individual differences in the weighting of early versus late processing stages. Such investigations could potentially uncover the neural basis for individual variations in pareidolia susceptibility and their relationship to broader aspects of visual processing.

Theoretical implications

These findings make substantial contributions to our understanding of visual perception and neural processing. The temporal dynamics of visual object recognition reveal a sophisticated system capable of maintaining multiple interpretations simultaneously. The relationship between neural processing stages and behavioral flexibility demonstrates how the brain can adaptively respond to different task demands while maintaining access to multiple levels of representation. The maintenance of multiple representations in visual perception suggests a more complex model of visual processing than previously considered, while the integration of bottom-up visual features with top-down behavioral goals provides insight into how the brain resolves perceptual ambiguity. Our results support a model where visual perception emerges from the dynamic interplay between multiple processing stages, with task demands modulating the relative contribution of each stage to behavioral outcomes. This framework provides a more nuanced understanding of how the brain maintains and utilizes multiple levels of visual information to support diverse perceptual judgments.

Data availability

Stimuli and behavioral data are available at https://doi.org/10.5281/zenodo.15833508. EEG data are available at https://doi.org/10.18112/openneuro.ds005642.v1.0.0.

Code availability

Code is available at https://doi.org/10.5281/zenodo.15833508.

References

1. Taubert, J., Van Belle, G., Vanduffel, W., Rossion, B. & Vogels, R. The effect of face inversion for neurons inside and outside fMRI-defined face-selective cortical regions. *J. Neurophysiol.* **113**, 1644–1655 (2015).
2. Quian Quiroga, R. et al. Single neuron responses underlying face recognition in the human midfusiform face-selective cortex. *Nat. Commun.* **14**, 5661 (2023).
3. Dy, T., Wa, F., Rb, T. & Ms, L. A cortical region consisting entirely of face-selective cells. *Science (New York, NY)* **311**, 670–674 (2006).
4. Kanwisher, N., Mcdermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
5. Rossion, B., Hanseeuw, B. & Dricot, L. Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain Cogn.* **79**, 138–157 (2012).
6. Puce, A., Allison, T., Asgari, M., Gore, J. C. & McCarthy, G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* **16**, 5205 (1996).
7. Taubert, J. et al. Clutter substantially reduces selectivity for peripheral faces in the macaque brain. *J. Neurosci.* **42**, 6739–6750 (2022).
8. Azadi, R., Lopez, E., Taubert, J., Patterson, A. & Afraz, A. The causal link between neural activity in inferotemporal cortex and free viewing eye movements. *J. Vis.* **23**, 5739 (2023).
9. Afraz, S.-R., Kiani, R. & Esteky, H. Microstimulation of inferotemporal cortex influences face categorization. *Nature* **442**, 692–695 (2006).
10. Sadagopan, S., Zarco, W. & Freiwald, W. A. A causal relationship between face-patch activity and face-detection behavior. *eLife* **6**, e18558 (2017).
11. Kanwisher, N. Domain specificity in face perception. *Nat. Neurosci.* **3**, 759–763 (2000).
12. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4–4 (2008).
13. Omer, Y., Sapir, R., Hatuka, Y. & Yovel, G. What Is a Face? Critical features for face detection. *Perception* **48**, 030100661983873 (2019).
14. Taubert, J., Wardle, S. G., Flessert, M., Leopold, D. A. & Ungerleider, L. G. Face Pareidolia in the Rhesus monkey. *Curr. Biol.* **27**, 2505–2509.e2 (2017).
15. Meng, M., Cherian, T., Singal, G. & Sinha, P. Lateralization of face processing in the human brain. *Proc. R. Soc. B: Biol. Sci.* **279**, 2052–2061 (2012).
16. Keys, R. T., Taubert, J. & Wardle, S. G. A visual search advantage for illusory faces in objects. *Atten. Percept. Psychophys.* **83**, 1942–1953 (2021).
17. Epihova, G., Cook, R. & Andrews, T. J. Recognition of pareidolic objects in developmental prosopagnosic and neurotypical individuals. *Cortex* **153**, 21–31 (2022).
18. Saurels, B. W., Peluso, N. & Taubert, J. A behavioral advantage for the face pareidolia illusion in peripheral vision. *Sci. Rep.* **14**, 10040 (2024).
19. Decramer, T., Premereur, E., Zhu, Q., Van Paesschen, W. & Van Loon, J. Single-unit recordings reveal the selectively of a human face area. *J. Neurosci.* **41**, 9340 (2021).
20. Rekow, D., Baudouin, J.-Y., Brochard, R., Rossion, B. & Leleu, A. Rapid neural categorization of facelike objects predicts the perceptual awareness of a face (face pareidolia). *Cognition* **222**, 105016 (2022).
21. Wardle, S. G., Taubert, J., Teichmann, L. & Baker, C. I. Rapid and dynamic processing of face pareidolia in the human brain. *Nat. Commun.* **11**, 4518 (2020).
22. Wardle, S. G., Paranjape, S., Taubert, J. & Baker, C. I. Illusory faces are more likely to be perceived as male than female. *Proc. Natl. Acad. Sci. USA* **119**, e2117413119 (2022).
23. Taubert, J., Wally, S. & Dixson, B. J. Preliminary evidence of an increased susceptibility to face pareidolia in postpartum women. *Biol. Lett.* **19**, 20230126 (2023).
24. Uchiyama, M. et al. Pareidolias: complex visual illusions in dementia with Lewy bodies. *Brain* **135**, 2458–2469 (2012).
25. Liu, J. et al. Seeing Jesus in toast: neural and behavioral correlates of face pareidolia. *Cortex* **53**, 60–77 (2014).
26. Thome, I. et al. "I spy with my little eye, something that is a face…": a brain network for illusory face detection. *Cereb. Cortex* **32**, 137–157 (2021).
27. Romagnano, V. et al. Dynamic brain communication underwriting face pareidolia. *Proc. Natl. Acad. Sci. USA* **121**, e2401196121 (2024).
28. Hebart, M. N., Zheng, C. Y., Pereira, F. & Baker, C. I. Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nat. Hum. Behav.* **4**, 1173–1185 (2020).
29. Grootswagers, T., Robinson, A. K., Shatek, S. M. & Carlson, T. A. Mapping the dynamics of visual feature coding: insights into perception and integration. *PLoS Comput. Biol.* **20**, e1011760 (2024).
30. Gupta, P. & Dobs, K. Human-like face pareidolia emerges in deep neural networks optimized for face and object recognition. *PLoS Comput Biol.* **21**, e1012751 (2025).
31. de Leeuw, J. R. jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res Methods* **47**, 1–12 (2015).
32. Peirce, J. et al. PsychoPy2: experiments in behavior made easy. *Behav. Res Methods* **51**, 195–203 (2019).
33. Oostenveld, R. & Praamstra, P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* **112**, 713–719 (2001).
34. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
35. Grootswagers, T., Robinson, A. K. & Carlson, T. A. The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage* **188**, gro (2019).
36. Robinson, A. K., Grootswagers, T. & Carlson, T. A. The influence of image masking on object representations during rapid serial visual presentation. *NeuroImage* **197**, 224–231 (2019).
37. Robinson, A. K., Quek, G. L. & Carlson, T. A. Visual representations: insights from neural decoding. *Annu. Rev. Vis. Sci.* **9**, 313–335 (2023).
38. Carlson, T. A., Grootswagers, T. & Robinson, A. K. *An Introduction to Time-Resolved Decoding Analysis for M/EEG* https://doi.org/10.7551/mitpress/11442.003.0075 (2020).
39. Oosterhof, N. N., Connolly, A. C. & Haxby, J. V. CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front. Neuroinform.* **10**, 27 (2016).
40. Nili, H. et al. A toolbox for representational similarity analysis. *PLoS Comput Biol* **10**, e1003553 (2014).
41. Dienes, Z. Bayesian versus orthodox statistics: which side are you on? *Perspect. Psychol. Sci.* **6**, 274–290 (2011).
42. Dienes, Z. How Bayes factors change scientific practice. *J. Math. Psychol.* **72**, 78–89 (2016).
43. Jeffreys, H. *Theory of Probability* (Oxford University Press, Oxford, 1961).
44. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D. & Iverson, G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16**, 225–237 (2009).
45. Wagenmakers, E. A practical solution to the pervasive problems of *p* values. *Psychon. Bull. Rev.* **14**, 779–804 (2007).
46. Morey, R. D. et al. *Package 'BayesFactor'* https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf (2018).
47. Wetzels, R. & Wagenmakers, E.-J. A default Bayesian hypothesis test for correlations and partial correlations. *Psychon. Bull. Rev.* **19**, 1057–1064 (2012).

48. Zellner, A. & Siow, A. Posterior odds ratios for selected regression hypotheses. *Trab. Estad. Investig. Oper.* **31**, 585–603 (1980).

49. Teichmann, L., Moerel, D., Baker, C. & Grootswagers, T. An empirically driven guide on using Bayes factors for M/EEG decoding. *Apert. Neuro* **2**, 1–10 (2022).

50. Wetzels, R. et al. Statistical evidence in experimental psychology: an empirical comparison using 855 *t* tests. *Perspect. Psychol. Sci.* **6**, 291–298 (2011).

51. Stuart, G., Saurels, B. W., Robinson, A. & Taubert, J. One object with two identities: the rapid detection of face pareidolia in face and food detection tasks. *J. Exp. Psychol.: Hum. Percept. Perform.* https://doi.org/10.31234/osf.io/5983j (2025).

52. Ritchie, J. B., Wardle, S. G., Vaziri-Pashkam, M., Kravitz, D. J. & Baker, C. I. Rethinking category-selectivity in human visual cortex. Preprint at https://doi.org/10.48550/arXiv.2411.08251 (2024).

53. Kanwisher, N. Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. USA* **107**, 11163–11170 (2010).

54. Sliwa, J. & Freiwald, W. A. A dedicated network for social interaction processing in the primate brain. *Science* **356**, 745–749 (2017).

Acknowledgements

Author contributions

Amanda K. Robinson: Conceptualization, methodology, software, formal analysis, resources, data curation, writing—original draft, writing—review and editing, visualization, supervision, project administration, funding acquisition. Greta Stuart: Methodology, software, investigation, and data curation. Sophia Shatek: Methodology, software, investigation, data curation, writing—review and editing. Adrian Herbert: Methodology, software, investigation, data curation, writing—review & editing. Jessica Taubert: Conceptualization, methodology, resources, writing—original draft, writing—review and editing, visualization, supervision, project administration, and funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information