communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-08497-8

Text-related functionality and dynamics of visual human pre-frontal activations revealed through neural network convergence

Check for updates

Adva Shoham [1] ✉, **Rotem Broday-Dvir** [2], **Itay Yaron** [3], **Galit Yovel** [1,3,4] **& Rafael Malach** [2,4] ✉

Human prefrontal areas show enhanced activations when individuals are presented with images, under diverse task conditions. However, the functional role of these increased activations remains a deeply debated question. Here we addressed this question by comparing, dynamically, the relational structure of prefrontal activations and both visual and textual-trained deep neural networks (DNNs) during a visual memorization task. We analyzed intra-cranial recordings, conducted for clinical purposes, while patients viewed and memorized images of familiar faces and places. Our results reveal that relational structures in the frontal cortex elicited during visual memorization were predicted by text and not visual DNNs. Importantly, the temporal dynamics of these correlations showed striking differences, with a rapid decline over time for the visual component, but persistent dynamics including a significant image offset response for the text component. The results point to a dynamic text-related function of prefrontal cortex during visual memorization in the human brain.

While a large body of human brain research has centered on characterizing the functional organization of the visual system located at the posterior part of the brain[1], a consistent finding in visual research reveals activation of pre-frontal cortex that is time-locked to visual images as well. These activations, although showing substantial report-related modulations (e.g., [2]), have also been demonstrated in relatively passive report-free situations such as movie watching[1,3–7] or viewing static stimuli[8–11], as well as during the encoding stage of image memorization tasks[12–14]. However, despite these consistent reports of frontal cortex activations to visual images under diverse task conditions, their functional role and information coded by these frontal areas have remained unclear. This question is of particular interest during image memorization, since, for example, it is likely that participants may be engaged in linguistic processing of the images during encoding. Yet, it is unknown whether these frontal activations reflect visual or semantic processing of information.

In recent years, the striking success of deep neural networks (DNNs) opened a new window for addressing the difficult problem of cortical functionality (see[15,16]). A particularly fruitful research direction has been the search for parallel relational structures in brains and artificial networks[17–22] (for review, see in ref. 23). Relational structures are a proposed representational scheme in which cognitive content, e.g., a visual image, is represented by the set of similarities and differences of this image from other visual images[24–27]. Such representations are particularly convenient when comparing different systems and modalities. In the domain of human perception, various recent studies have shown that such DNN-derived relational structures account for a significant proportion of variance in human representations of faces and objects, outperforming any previous computational models[23,28–31]. Given the finding of significant parallels between artificial networks and visual representations at the posterior cortex, it is of interest to search for such parallels in the visually active sites found in the human frontal lobe.

Previous studies that used DNNs to predict brain activity to visual images primarily used DNNs that were trained on visual inputs[29,32,33]. In a recent study, Shoham et al. [34] proposed a novel approach, involving models trained on different types of information (visual or textual) to test whether they explain unique perceptual and semantic components in the representation of familiar visual stimuli in human perception and memory. This approach is ideal to address the question of whether neural activations to visual images reflect the processing of visual or linguistic information. Specifically, we aimed to discern whether the response of

[1]School of Psychological Sciences, Tel Aviv University, Tel Aviv, Israel. [2]Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel. [3]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel. [4]These authors jointly supervised this work: Galit Yovel, Rafael Malach.
✉e-mail: advashoham@mail.tau.ac.il; advash92@gmail.com; rafi.malach@gmail.com

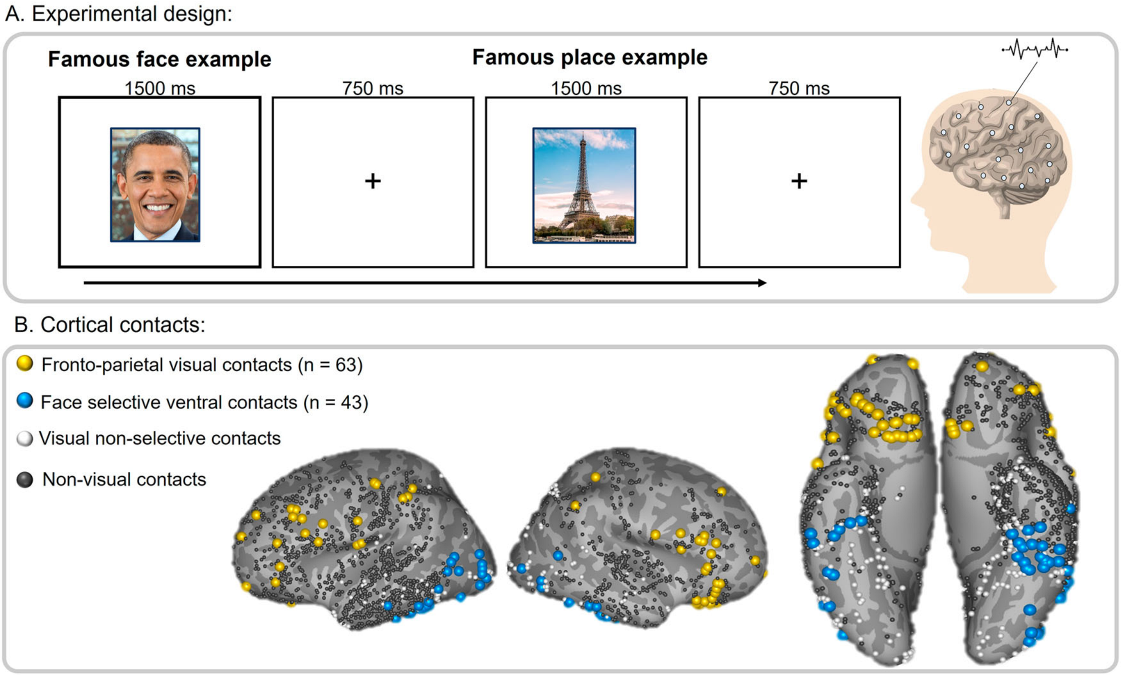

Fig. 1 | Experimental design and general methods. **a** Experimental design: participants viewed 28 images of familiar faces or places while their brain activity was recorded via intracranial EEG. **b** Contact locations: yellow - Fronto-parietal visual contacts, blue— Face-selective ventral contacts, white—visual non-selective contacts, gray- non-visual contacts. The original images used in the experiment are not copyrighted, images in the figure were replaced by licensed images with a similar appearance. Credits: Barack Obama image is from Whitehouse.gov under a Creative Commons license CC BY 3.0, Eiffel Tower image is from Freepik (https://freepick.com).

the frontal lobe to visual stimuli during an image memorization task is primarily representing visual or textual information. To achieve this, we employed visual (ImageNet-trained VGG-16) and text (GPT2) DNNs to model brain activity. Recent findings showed that the multi-modal, image-language DNN, CLIP (Contrastive learning image pre-training) also accounts for a significant proportion of variance beyond the unimodal visual and semantic contributions in behavioral[34] and neural representations[35–37]. Therefore, in addition to standard visual and language DNNs, we also extracted visual and textual embeddings from CLIP image and text encoders.

To quantify the contributions of visual and textual information to activations driven by memorization of visual stimuli in the frontal lobe, we used intra-cranially prerecorded iEEG data in which patients were presented with images of famous individuals and places and were asked to visually inspect the images and memorize the visual details in each image for a later recognition test (specifically noting their visual attributes such as color and shape) (see Fig. 1). This dataset was previously employed in exploring questions regarding hippocampal and high-order visual cortex activations during memory recall[38,39] and the stability of relational structures in high-order visual cortex during viewing of sustained visual stimuli[8]. Here, we purposely focused on visually-responsive contacts in the frontal cortex that were not specifically examined priorly, to explore the multivariate nature of activations in visually responsive contacts located in the frontal cortex during memorization. The high resolution of the iEEG brain recordings and their comparison to DNN representations enabled us to follow the precise dynamics of these frontal processes while participants viewed and memorized images, and investigate whether they are mainly driven by visual or textual information. Our findings showed that the activity of visually responsive fronto-parietal contacts was best predicted by text-based rather than visual-based networks. These results indicate that the response of front-parietal cortex to visual images reflects semantic rather than pure visual processing.

Results

Experimental design and intra-cranial contacts information

Our study included intra-cranial recordings, conducted for clinical purposes in **13** patients (10 females, mean age 34.7 ± 9.6) and 2571 recording sites. The participants were presented with images of famous individuals and places. Briefly, participants viewed 28 different images, 14 familiar places and 14 familiar faces, divided into two experimental runs. Each image was presented four times during the run, in a pseudo-random order, ensuring no image was repeated twice consecutively. The images were presented for a sustained time duration of 1500 ms, with 750 ms inter-stimulus fixation intervals. No participants were excluded based on performance (for further details see methods and[38,39]).

Figure 1 depicts the experimental set-up and recording sites that were examined in the present study. We first identified the visually responsive contacts ($n = 377$, out of the 2571 total contacts) and then divided them, based on their anatomical locations (ventral, fronto-parietal) and functional characteristics, into the different regions of interest (ROIs) (see methods of[39] for details).

Extraction of visual and textual representations from DNNs

Visual representations of the 28 images were extracted from the final fully connected layer of an ImageNet-trained DNN (VGG-16) and from the final layer of the visual encoder of CLIP. Textual representations of the same familiar stimuli were the embeddings of the first paragraph of their Wikipedia description based on the final layer of GPT2, and the embeddings of their names, based on the textual encoder of CLIP (Supplementary Fig. 10 shows that CLIP's name-based representations carry relevant semantic information. See Methods section for details). We then generated representational dissimilarity matrices (RDMs) by computing the cosine distance between the embeddings of the 28 images/textual descriptions for each DNN (see Fig. 2A). Resulting in four different RDMs, two visual-based RDMs (VGG, CLIP-Image) and two text-based RDMs (CLIP-Text, GPT). The correlations between the different DNNs are presented in Fig. 2B.

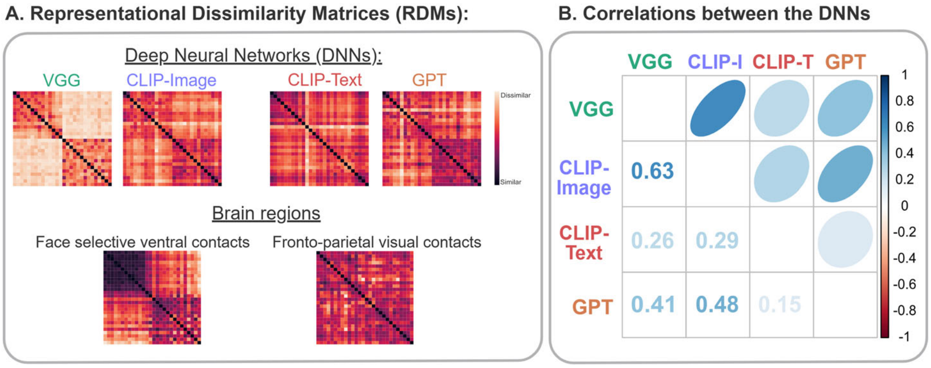

Fig. 2 | Representational dissimilarity matrices. **A** Representational Dissimilarity Matrices (RDMs) of the 28 stimuli. RDMs of the embeddings of the images based on visual DNNs (VGG, CLIP-image) and of their textual description based on textual DNNs (CLIP-text, GPT). RDMs of the neural response to the images in face-selective ventral and fronto-parietal visual contacts, based on multivariate responses across contacts. **B** The correlations between the RDMs of the visual and textual DNNs. The thinner the ellipsoids, the greater the correlation.

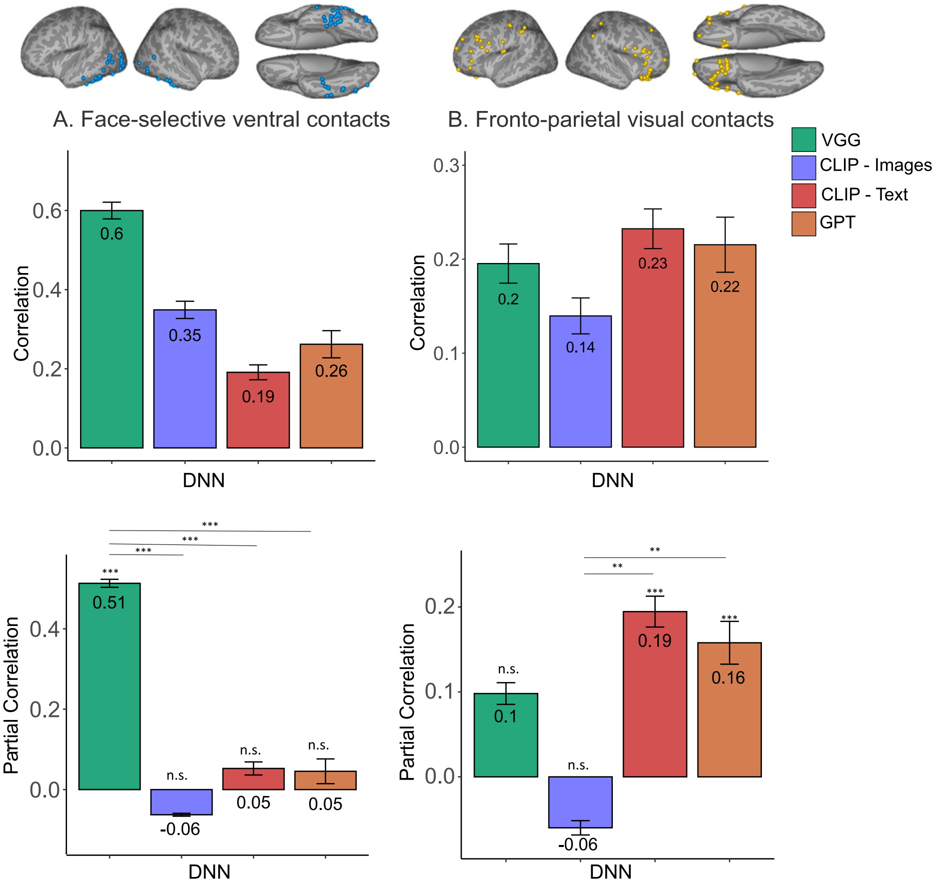

Fig. 3 | Correlations (Top) and Partial correlations (bottom) of image and text DNNs with Face-selective ventral and Frontal-parietal visual contacts. Zero order correlations are presented in top panels, and partial correlations with each of the DNNs, when the other three DNNs are held constant, are presented in bottom panels. **A** Face-selective occipito-temporal contacts ($n = 43$), Error bars indicate leave one participant out procedure s.e.m. **B** Fronto-parietal visual contacts ($n = 63$), Error bars indicate leave one participant out procedure s.e.m. All p values were derived from a pair-images permutation test (10,000 permutations). Reported p values are FDR corrected in each ROI separately. Note the clear shift in bias from visual-based to text-based preference from occipito-temporal to frontal-parietal ROIs.
$**p_{FDR} < 0.01$, $***p_{FDR} < 0.001$.

Fronto-parietal visual contacts are correlated with text-based rather than visual-based DNNs

To explore the potential parallels between the activations in visually responsive contacts located in the fronto-parietal cortex and DNNs, we compared their relational structures as revealed through representational dissimilarity matrices (RDMs, see in ref. 28). For each stimulus, we calculated the average response in each contact within a range of 0.1–0.4 s (100–400 ms) from the moment the stimulus was presented. We mainly focused on this time window as it encompasses the peak of the response, but we explored the response averaged across other time windows as well (time windows at 0.4–0.7 s, 0.7–1.0 s, 1.0–1.3 s, and 1.5–1.8 s). We defined the contacts' mean response (averaged across 4 repetitions) population vectors as the stimulus activation patterns in each ROI and measured the Pearson correlation between each two stimuli activation patterns. One minus correlation represented the dissimilarity score between the two stimuli (lower = more similar) in this ROI. This provided us with the RDM for each ROI (see Fig. 2A).

Next, we compared the relational-structure similarities between the fronto-parietal contacts and the output layers of four different networks: two networks were text-based (GPT and CLIP-Text), and two networks were image-based (CLIP-image and VGG). For further details about the different DNNs, see methods. The correspondence between the patients' fronto-parietal ROI RDMs (or the face-selective ventral ROI) and the artificial networks' layers was computed using Pearson correlations: We computed the correlation between the fronto-parietal ROI RDM with each network RDM (Fig. 3 top-right panel). All four networks were correlated with the response of the fronto-parietal cortex contacts to the images. However, our aim in this analysis was to examine the unique contribution of each DNN network to the overall variance. To test this unique similarity between each network representation with the fronto-parietal cortex we calculated the partial correlations between each DNN and the patients' RDM, when all other DNNs are held constant. To test for statistical significance, we computed the same correlations using a permutation test (shuffled pairs labels, see methods for further details) and bootstrap analyses (leaving one participant/contact/stimuli/repetition out). We tested both the significance of the predictors' partial correlations individually and the difference between each partial correlation in each ROI. P-values were FDR corrected for each ROI and each analysis (individual predictors and differences) separately.

Figure 3 (bottom panels) depicts the partial correlations found between each of the different DNNs (warm colors depict the text-related while cold colors the visual-related ones), when the other three DNNs are held constant. The figure shows the occipito-temporal, high-order, visual face-related ventral ROI (Panel A bottom) and the fronto-parietal visual contacts (Panel B bottom) in the 0.1–0.4 s time window. As can be seen, in the Fronto-Parietal ROI, only the text-related DNNs displayed significant partial correlations to the patients' RDMs. The opposite selectivity was found in the face-selective ventral contacts in the visual cortex. Here, only the visual-related VGG DNN showed a significant partial correlation to the Face-selective ventral contacts (note that for zero-order correlation (top panels), both VGG and CLIP-Image are highly correlated with the ventral-visual contacts. The reason the partial correlation with CLIP-Image is low is because it does not contribute unique variance beyond the information that VGG and the text networks contribute).

As shown in Fig. 3B top panel, all four networks were correlated with the Fronto-Parietal ROI. To better understand whether the visual networks (whose correlation dropped in the partial correlation analysis) account for the same information as the text networks, we added a category RDM as a fifth predictor in the partial correlation analysis. The category RDM was implemented by assigning similarity scores of 0 for same-category pairs (face-face or place-place) and 1 for different-category pairs (face-place). We then calculated the partial correlations between each predictor and the Fronto-Parietal ROI. When adding the category RDM as a predictor, the visual networks showed very low and non-significant partial correlations. In contrast, adding the category RDM did not alter the significant association between the Fronto-Parietal ROI and the text-based DNNs (see Supplementary Fig. 1). These findings suggest that the association of the text-based networks with the Fronto-Parietal ROI is not category-based.

Note, that although the fronto-parietal visually responsive contacts show weaker correlations with the DNNs compared to the ventral contacts, they also show substantially noisier responses as manifested in the lower limit of noise ceiling measure (noise ceiling of 0.30 compared with 0.77, respectively). Therefore, the correlation with the fronto-parietal contacts and the correlations with the ventral contacts should be each evaluated relative to their own noise ceiling (see method section for noise ceiling calculation).

Since each patient was presented with the same visual images four times, we wanted to ensure that the pattern observed in Fig. 3B (bottom) was not primarily driven by a single repetition. To investigate this, we calculated the partial correlations when excluding one repetition each time. The dominance of the text-based DNNs remained consistent across all 'leave-one-repetition-out' iterations (see Supplementary Fig. 2).

Finally, given that the stimuli set in this study included both places and faces, we employed a second control ROI that was not only face-selective but rather included contacts in high-order visual cortex that showed significant preference for a subset of the stimuli (could be faces, places, or a mix of both-Content Selective ROI, see Supplementary Material). The pattern observed in this ROI was similar to the results found for faces-selective ventral contacts (for details, see Supplementary Material and Supplementary Fig. 3).

Sustained textual contribution with a stimulus-offset response in fronto-parietal visually responsive contacts

Given that CLIP-Text showed the strongest partial correlation with the fronto-parietal visually responsive contacts, we proceeded to investigate this brain-network correlation across different time windows (Fig. 4). For comparative analysis, we examined CLIP-Text correlations within the same time windows with face-selective ventral contacts as well as the correlations of the ventral and fronto-parietal contacts with the visual DNN (VGG) (note that analysis shown in Fig. 4 displays the correlations and not the partial correlations that are shown in Fig. 3 bottom panels). To assess the significance of differences in correlations across time windows, we used a permutation test (see methods for further details). When exploring CLIP-Text and fronto-parietal correlations (Fig. 4, bottom-right panel), they were all as high as the correlation of the first time window (0.1–0.4 s), besides the 1.0–1.3 s, which showed no correlation. This might suggest a relatively persistent processing pattern during the stimuli presentation for about 1 s.

Interestingly, a similar pattern seems to reappear at the stimulus offset (1.5 s). This was not the case for VGG correlations with the fronto-parietal visual contacts (Fig. 4 top-right panel), which showed a steep decline after the first time window. Nor is the case for the face-selective ventral contacts that showed gradual decline in VGG correlations (Fig. 4 top-left panel) and variable CLIP-Text correlations with time (Fig. 4 bottom-left panel). See Supplementary Fig. 4 for a similar analysis with GPT. See also Supplementary Fig. 5 for an exploratory analysis of the unique contribution of visual and semantic information to the representation of the fronto-parietal cortex in each time-point, by calculating the partial correlations between VGG and CLIP-Text and the patients' RDM.

A monotonic increase in the correlations with fronto-parietal visual contacts across the layers of CLIP-Text

To further explore the initial-time window (0.1–0.4 s), we asked whether the frontal correlation is merely due to the inner structure of the DNNs, regardless of its specific input contents. We also asked how the fronto-parietal correlation is manifested across the different layers of the DNN. Figure 5 shows that the correlation to CLIP-Text based on shuffled text input is drastically reduced compared to the original text (the "output" layer was used in the previous analyses; see Supplementary Fig. 11 for the same comparison between GPT's embeddings based on the original and shuffled text, which shows similar results). We also extracted the RDM of each layer of the CLIP-Text network (see methods for further details). Examining the evolution of the correlation to the Fronto-parietal contacts' RDM through the DNN layers revealed a clear monotonic increase towards higher order layers, reaching the highest values at the three top layers (Residual attention blocks (RAB) 11,12, and output-13), indicating a specific, functional similarity between the fronto-parietal contacts' responses and the top CLIP-Text layers. Also note the striking difference between the RDM correlations of the correct and shuffled texts. For the correlation of the fronto-parietal contacts with VGG layers, which shows a different pattern, lacking a monotonic increase, see Supplementary Fig. 6.

To rule out the possibility that these correlations were a result of an outlier patient, contact, or stimulus response, we examined how robust the correlation was to CLIP-Text when using only partial data in the analysis. These analyses and results are reported in the Supplementary material. As can be seen in Supplementary Fig. 7, the results remained fairly robust to different bootstrap analyses, arguing against an outlier-driven effect.

Correlations of CLIP-Text with fronto-parietal visually responsive contacts are maximal for the first image repetition

Since the same visual images were shown four times to each patient, it was of interest to examine how such repetitions affected the observed correlation to the CLIP-Text network. To that end, separate RDMs were constructed for each repetition of the stimuli. Supplementary Fig. 8 depicts the RDM correlations across CLIP-Text layers for each of the repetitions. Results show that the correlation levels changed in a complex manner across repetitions. With the novel images (first appearance) failing to show any correlation, and a significant correlation appearing for the first repetition followed by a progressive decline in the second and third ones.

Finally, to examine whether the correlation to CLIP-Text changes across different fronto-parietal subdivisions, the contacts were divided into two major anatomical subdivisions, orbito-frontal and lateral-frontal contacts (see further details in the Supplementary file). As can be seen in

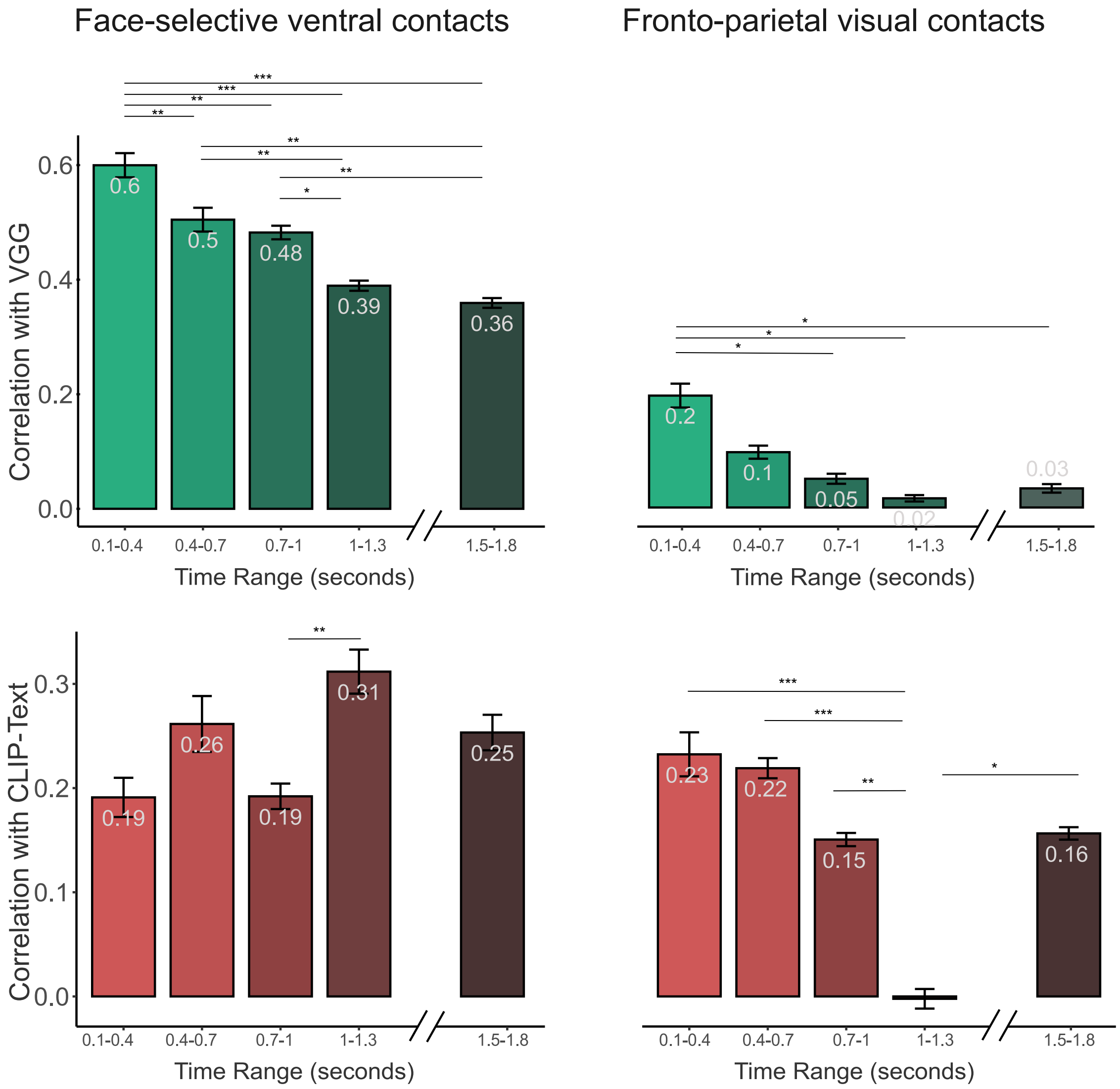

Fig. 4 | Correlations between VGG or CLIP-Text with Face selective ventral and Frontal-parietal visual contacts at different time windows. Top bar plots (green bars) show the correlations with VGG in face-selective ventral contacts (left), and fronto-parietal visual contacts (right). Bottom bar plots (red bars) show the correlations with CLIP-Text in face-selective ventral contacts (left), and fronto-parietal visual contacts (right). Note the strikingly different dynamics in fronto-parietal contacts between the visual component that declined rapidly and the text component with persisted followed by an offset response. Error bars indicate leave one participant out procedure s.e.m $*p_{FDR} < 0.05$, $**p_{FDR} < 0.01$, $***p_{FDR} < 0.001$.

Supplementary Fig. 9, the overall pattern of correlations across anatomical regions in the frontal cortex remained largely unchanged, showing a gradual increase in correlation when moving from low-level to high-level network layers.

Semantic similarity, rather than visual similarity ratings, uniquely predicts the fronto-parietal neural response

To further examine if fronto-parietal processes during the memorization of visual images are mainly driven by visual or semantic information, we tested the correlation of this ROI with visual and semantic human similarity ratings. We asked a different group of participants to rate the visual or semantic similarity of all possible pairs among the 28 stimuli. One group ($n = 20$) rated visual similarity based on the appearance of image pairs. The other group ($n = 20$) rated semantic similarity presented with the names of the stimuli (e.g., the identity or place depicted), based on their conceptual knowledge (see Fig. 6 and Methods for details). We created RDMs based on the average similarity ratings across participants in the visual and semantic tasks and calculated their correlations and partial correlations with the fronto-parietal contacts. To test for the statistical significance of each correlation (and partial correlation), we computed the same correlations using a permutation test (shuffled pairs labels, see methods for further details). As can be seen in Fig. 6, in the Fronto-Parietal ROI, only the semantic ratings displayed a significant partial correlation. The opposite pattern of correlations was found in the face-selective ventral contacts in the visual cortex. These findings further highlight the association of the Fronto-Parietal neural responses to visual images during memorization with semantic information.

Discussion

In this study, we used text-based and image-based DNNs to investigate their similarity to brain activity in visually responsive sites, while participants viewed and memorized images of famous faces and places. Our analysis revealed that visually responsive fronto-parietal contacts' activations were best predicted by text-based networks, suggesting semantic rather than visual-related processing during image memorization.

Frontal lobe representations in visual memorization processing

In agreement with a number of prior studies based on both fMRI and intracranial brain recordings that revealed that the frontal lobes are consistently activated to the presentation of visual images during

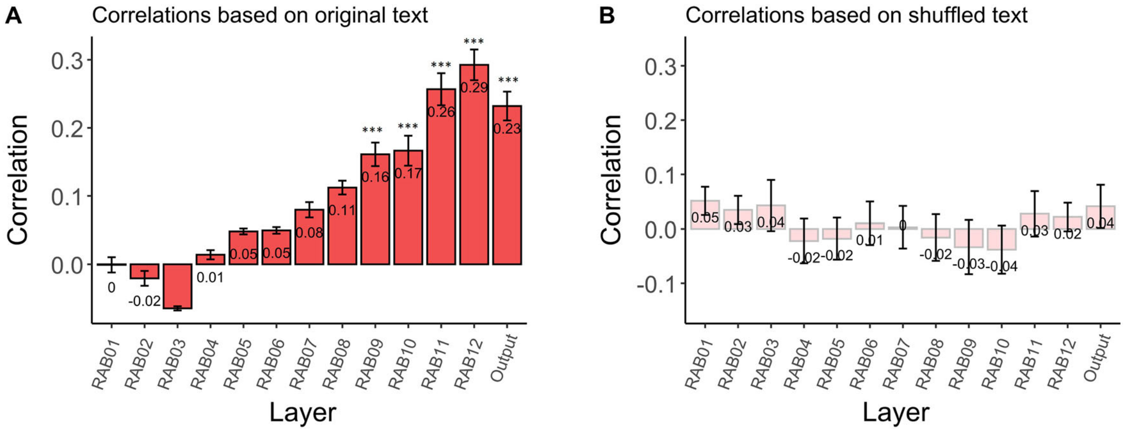

Fig. 5 | CLIP-Text different layers' correlations with Fronto-parietal visual contacts when embeddings were based on the original text or on shuffled text. The correlations with the fronto-parietal visual contacts ($n = 63$) in the first time-window (0.1–0.4 s) with CLIP-Text different layers based on the original text or shuffled text. **A** The correlations with CLIP-Text representations based on the original text, Error bars indicate leave one participant out procedure s.e.m. **B** The average correlation with CLIP-Text different layers based on six different variations of shuffled original text. Error bars indicate s.e.m. All p values were derived from a pair-image permutation test (10,000 permutations). $* * p_{FDR} < 0.01$, $* * * p_{FDR} < 0.001$.

perceptual[2,3,8,9,11,40,41] and memorization[12–14] tasks. Our study also reveals image-responsive frontal activations. Importantly, in the present study, the patients were instructed to view and memorize the visual details of the images but were not instructed to report seeing the visual images- ruling out the attribution of the frontal activations to explicit reporting[5].

Though previous studies showed frontal lobe activations while viewing[3,8,9,40,42] or memorizing[12–14] visual stimuli, the specific role of these frontal lobe responses to images is still unclear. Some research highlights the non-visual nature of frontal lobe tasks, with the highest activations during reporting and motor tasks[2,5,43,44], and lesion cases challenge the idea of the prefrontal cortex's essential role in perception[45,46]. It is important to note that identifying the function of cortical responses is a fundamental challenge in all human brain studies, particularly since causal manipulations are rare and necessarily associated with poorly controlled clinical cases. We and others have recently proposed[15,34,47,48] that insight into the functionality of cortical responses can be gained by searching for "convergent evolution"—i.e., parallels that are found between artificial networks and brain systems[16,28,49]. Here, we adopted this strategy to gain insights into the enigma of frontal responses to visual images during a memorization task. By searching for correlations between relational structures, revealed through RDMs, in frontal cortex and visual and textual deep learning networks, we have uncovered a preference for text-related networks. The correlations to the image-based artificial networks although positive, were far weaker and did not account significantly for the fronto-parietal response beyond the textual-based networks.

The fact that the more specific correlation was found to a text-based network rather than a similar visually related ones (CLIP-image) is intriguing, especially considering that the patients' RDMs were based on visual responses, during a task targeting memorizing specifically the visual details of the image, and that visual-related DNNs were correlated with visual contacts in high level visual cortex. This finding suggests that the frontal responses to memorized visual stimuli may be more closely tied to linguistic analysis driven by visual contents rather than visual perception per se (see in ref. 50 for the opposite view). Notably, robust visual responses associated with non-perceptual functions are well-documented in the human brain. Particularly striking cases are visual activations found in the hippocampus associated with conceptual representations[51,52] and visual activations in human high-order somatosensory cortex[53,54]. The visual activations in this study appear to be more linked to linguistic rather than perceptual processing functions and thus are compatible with prior studies supporting a role of the frontal lobes in language and thought processing[55]. Furthermore, while the participants in this study performed a visual memorization task, the frontal lobe responses appeared to represent the linguistic over the perceptual information during this task.

Participants in the current study were instructed to memorize specific visual details of the images during the memorization task. In a recent paper[56], using the same iEEG data, we dissociated between two types of textual descriptions of images, *visual text*, which describes visual content of the image, similar to the information participants encoded in this task, and semantic text, which is the knowledge that we have about and image that is independent of its appearance such as their name or location. Our findings show that artificial language model representations of the visual text, but not the semantic text, were correlated with the response of the visual cortex to the images. Interestingly, the representation of the semantic text, but not the visual text, was significantly correlated with the neural activation of the fronto-parietal cortex. These findings indicate that it is not the encoding of the visual details of the images that underlies the activation of the frontal lobe to images, but their semantic linguistic information.

The temporal dynamics of prefrontal and posterior correlations

Comparing the temporal dynamics of correlations between the RDMs of DNNs and brain activations revealed different patterns of results in posterior compared to fronto-parietal ROIs. In the posterior face selective ventral ROI both image-based (VGG) and text-based (CLIP-Text) networks showed consistent correlations across time windows, with higher correlations with the image-trained DNN (VGG). Conversely, activations in the fronto-parietal ROI were correlated with image-trained DNN (VGG) only in the earliest time window (0.1–0.4 s), in sharp contrast to the correlations profile of CLIP-Text, which were found to be significant in all but one of the time windows (1–1.3 s). These results echo previous findings showing persistent representations of visual contents in posterior regions[8,9,40], and add novel insight regarding the dynamics of prefrontal activations in two ways: First, the obtained results point to a differential dynamics for the transient, visual-related aspects of the fronto-parietal activations vs the persistent text-related ones. Second, by focusing on the text-based component of the fronto-parietal activation, our results demonstrate an intriguing offset response. It is worth mentioning that whereas both CLIP-Text and GPT were similarly correlated with fronto-parietal contacts during the early latency, they exhibit distinct patterns of correlation when looking at later time windows (Fig. 4 and Supplementary Fig. 4). These variations might be due to their different input (name in CLIP-Text or paragraph from Wikipedia in GPT) or the way they were trained. CLIP-Text is a multi-modal visual language DNN, while GPT is a unimodal language DNN. Further research is needed to account for these differences.

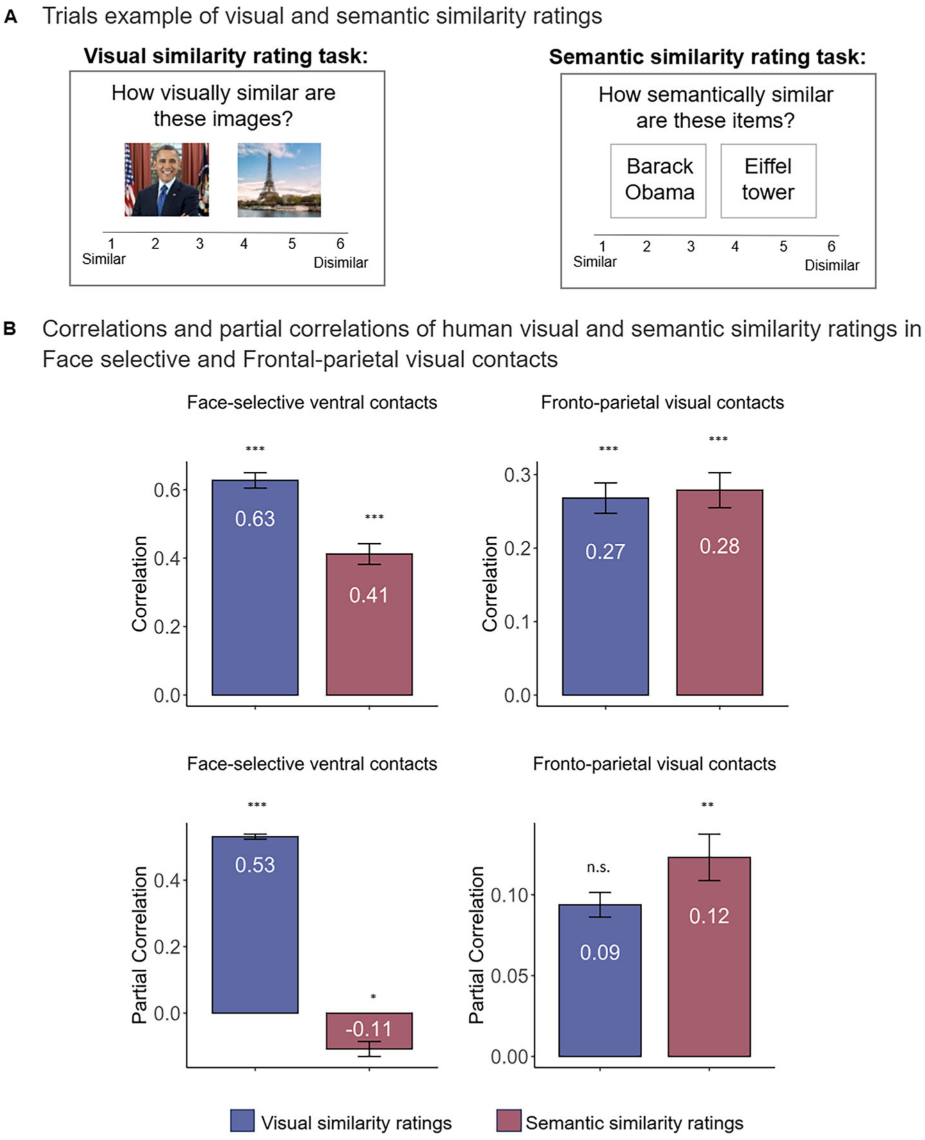

Fig. 6 | Correlations (Top) and Partial correlations (bottom) of human visual and semantic similarity ratings with Face-selective ventral and Frontal-parietal visual contacts. Zero-order correlations are presented in top panels, and partial correlations between each similarity task RDM (averaged across participants; $n = 20$ in each task), when the other task RDM is held constant, are presented in bottom panels. **A** Face-selective occipito-temporal contacts ($n = 43$), Error bars indicate leave one participant out procedure s.e.m. **B** Fronto-parietal visual contacts ($n = 63$), Error bars indicate leave one participant out procedure s.e.m. All p values were derived from a pair-image permutation test (10000 permutations). Reported p values are FDR corrected in each ROI separately. $* * p_{FDR} < 0.01$, $* * * p_{FDR} < 0.001$. The original images used in the experiment are not copyrighted, images in the figure were replaced by licensed images with a similar appearance. Credits: Barack Obama image is from Whitehouse.gov under a Creative Commons license CC BY 3.0, Eiffel Tower image is from Freepik (https://freepick.com).

Dynamic modulation of correlations by stimulus repetition

The experimental design in the present study included four repetitions of each of the stimuli presented. The temporal distance between repetitions was randomized across stimuli- however, grouping the repetitions according to their sequential repetition order revealed a complex dynamic of the correlation magnitude to the CLIP-Text network. Thus, one could discern two, not necessarily linked, changes (see Supplementary Fig. 8). The first was a rapid increase from essentially no correlation in the first, novel image presentation to the first time the images were repeated. This suggests that the fronto-parietal cortex function that is correlated to the CLIP-Text network may not be an encoding function, but rather related to automatic short-term memory or recall. It is interesting to note that a similar discrepancy between novel and repeated presentations was observed in the hippocampal formation[38] where content-selectivity appeared only on a second presentation, while the first response was a more generalized novelty signal. It is intriguing to speculate that this novelty effect may have been dominant also in the pre-frontal recordings and masked the potential correlation with the CLIP-Text network. The second effect was a gradual decline in the correlation at further repetitions. This type of gradual decline, termed adaptation or repetitions suppression, has been ubiquitously observed in the visual system[8,57–60]. It is interesting to note that what appears as a repetition-suppression effect can be revealed through the window of brain-artificial network correlations.

Potential insight into theories of consciousness

There is an ongoing debate concerning the functional role of visually driven prefrontal activations in human consciousness. Of particular significance, the hypothetical function assigned to pre-frontal activations and their temporal dynamics has played a major role in global theories of consciousness[61]. For example, "localist" theories (e.g., [24,62,63]), positing that different contents of awareness should be linked to information in local cortical regions, will argue that the visual pre-frontal regions are likely linked to non-perceptual aspects, typically attributed to the frontal lobe- such as language and decision making. In contrast, "globalist" theories- such as Global Neuronal Workspace[64,65], or High Order Theories[66,67] will attribute to frontal activations a central role in enabling visual perceptual awareness, and even representing the contents of conscious perception according to some views. Moreover, according to the *Global Neuronal* Workspace theory, prefrontal activations are hypothesized to be transient, showing at both the appearance and disappearance of perceived stimuli, reflecting the updating of consciously perceived contents[40,61] (for a recent review of this debate see in refs. 43,68). Thus, highlighting the potential functional role and dynamics of visual activation in human prefrontal cortex, beyond its immediate interest in understanding the functionality of the frontal lobes, has theoretical consequences. Our results demonstrate the intriguing offset response when focusing on the text-based component of the frontal activation is predicted, albeit for visual processing under relatively passive viewing conditions (up to

300 ms after stimulus offset), by the Global Neuronal Workspace theory[40]. However, our findings highlight the non-visual nature of processing of the frontal lobe, arguing against global theories of consciousness which posit that visual processing necessitates a spread to fronto-parietal cortex during conscious vision. However, it should be noted that our finding might be specific to the visual memorization task and therefore might not generalize to conscious perception per se.

Using deep neural networks to test the contribution of different types of information to neural representations

Current DNNs, such as the transformer models used here, were not originally designed to model the architecture of the brain, and their application in neuroscience must be approached with caution. These models are best understood as provisional starting points for generating and testing hypotheses about neural function, rather than as direct explanations. In line with this view, our use of different DNN architectures was not intended to imply that any one model offers a complete account of neural processing. Rather, our goal was to leverage their divergent training objectives (e.g., vision vs. language) to disentangle the types of representational content present in different brain regions. By contrasting the representational structures of these models, we could isolate whether specific brain areas, particularly in the fronto-parietal cortex, encode more visual or semantic information. Therefore, while these models are limited as biological accounts, their differences serve a valuable functional purpose in our study. The observed pattern of correlations, especially the alignment with text-based models in higher-level brain regions, provides empirical evidence for the presence of text-related (semantic) representations in these regions.

Limitations

Although we analyzed only contacts driven by the presentation of visual images, we cannot conclusively attribute our results to perception or memorization alone. The semantic representations we identified in the fronto-parietal cortex emerged at image presentation and were time-locked to the images (0–400 ms, Figs. 3, 5 and 6), suggesting they may reflect responses to the images themselves. However, since participants were instructed to memorize the images for a later memory task, these activations could also reflect rapid aspects of the memorization process. It should be noted that the present study did not attempt to disentangle the perceptual from the memorizing aspects of the task to the extent that this is at all possible[69]. Future studies should investigate whether the fronto-parietal cortex represents images differently during e.g., passive viewing versus memorization.

Examining the magnitude of the correlation between the RDMs derived from fronto-parietal activations and the CLIP-Text layers, even when specifically focusing on the optimal conditions, failed to reveal correlation levels higher than 0.3. This is a significantly lower value compared to magnitude of the brain to network correlation values found in high order visual areas—which was substantially higher, ranging from 0.4 to 0.6[28] (and see VGG correlations in Fig. 4). Therefore, it could be argued that the differential correlation pattern we found in fronto-parietal and posterior areas merely reflects a coarser grained representation in fronto-parietal cortex. However, the finding that text-models show higher correlation in fronto-parietal compared to posterior visual areas on the one hand, and the discovery of different temporal dynamics of the correlations in fronto-parietal vs posterior regions—argue against such possibility. Moreover, while the fronto-parietal correlation of 0.2–0.3 was lower than the visual correlation (0.6)—it should be noted that such correlations are consistent with a number of previous studies showing similarly lower, yet significant, correlations[70–77]. A number of sources, not necessarily mutually exclusive, could account for such relatively low correlation values. The most obvious factor is the relatively low activation level of the fronto-parietal contacts, especially relative to the ongoing noise. This weak activation and low SNR is typical in report-free paradigms such as the one adopted in the present study (see also in refs. 2,8), and is reflected in the lower noise ceiling. Whether this effect is simply due to smaller signal-to-noise ratios in the weakly responding contacts or whether it reflects a less selective signal in these contacts remains to be studied.

A second possibility is that, as indeed suggested by the preferential correlation to the text-related DNNs, the true function of the pre-frontal contacts is not perceptual, but semantics-related. Since the task of the patients did not involve explicit semantic or textual aspects, the visual activation may have merely occurred as an automatic, auxiliary by-product of the visual presentation and memorization of a familiar image. Hence, the obtained frontal activation may be less precisely time-locked to the presentation of images, leading to noisier, less consistent responses compared to the posterior perceptual activation. Resolving this issue will necessitate additional experiments in which semantics and decision-making tasks will be manipulated in addition to the visual images.

Another limitation to note is that we used a relatively small number of stimuli (*n* = 28), whereas recent fMRI studies have explored neural representations with a much larger set of stimuli[35,36]. iEEG is a superior method compared to non-invasive techniques such as fMRI or scalp EEG, as it combines high temporal resolution, allowing us to explore dynamic changes in correlations, with high spatial resolution, and is more directly related to the underlying neuronal activity. However, a major limitation of this unique method is that, due to patient fatigue, an extensive study including many images and categories is not feasible. Previous studies explored brain representation using different recording methods, have also used such limited set size of stimuli[28,70–72,76–78].

Finally, one cannot rule out the possibility that more advanced network structures may produce higher and more consistent activation levels. The recent explosive growth in the variety and power of language models will offer rich ground for future explorations in this domain.

To conclude, our findings suggest a closer association between visually responsive fronto-parietal regions and text-based rather than visually related DNNs. The temporal dynamics of this text-based association were observed in different latencies of the stimuli presentations with a drop and a recovery after the stimulus offset. Overall, our findings highlight the non-visual processing of the frontal lobe during the visual memorization of images.

Methods

Human iEEG data

We describe here the participants tasks and stimuli as well as ROI definitions. Additional details can be found in ref. 8.

Participants. Electrophysiological data were acquired from 13 patients (10 females, mean age 34.7 ± 9.6) via intracranial iEEG recordings. The recordings were obtained while the patients underwent pre-surgical evaluation for drug-resistant epilepsy at North Shore University Hospital in NY. As part of the clinical assessment, subdural or depth contacts were implanted in all patients. The study followed the latest version of Declaration of Helsinki, and all patients provided a fully informed consent to participate, including consent to publish the results, according to the US National Institute of Health guidelines. The Feinstein Institute for Medical Research's institutional review board monitored the study, and all ethical regulations relevant to human research participants were followed. Notably, no clinical seizures occurred during the experiment.

Experimental task and stimuli. The experiment comprised two runs, each starting with a 200-s resting phase with closed eyes. Subsequently, participants viewed 14 different images in each session (7 images of each category—famous faces or famous places). In total, 28 stimuli were used, 14 in each run (a similar stimuli set size was used previously in studies that measured neural representations of visual stimuli[28,70–72,76–78]). Each image was presented four times (for 1500 ms each time, with 750 ms inter-stimulus intervals) in a semi-random order to avoid consecutive repeats. Participants were instructed to look carefully at the pictures and note, specifically, the visual rather than semantic details. They were informed that they would later be required to recall the images and describe their visual features in detail, rather than merely identifying them. Consequently, they were instructed to memorize the details of the images. Participants performed well, and no participants were excluded based on performance (Further details can be found in ref. 39).

Definition and grouping of visually responsive contacts. Our study comprised of 2571 recording sites (contacts). However, the analyses focused only on visually responsive contacts, subdivided into different ROIs. Visual responsivity was defined as electrodes displaying statistically significant activations in response to stimuli presentations. To do so, the data from all contacts (visual and non-visual) was preprocessed and transformed to the high-frequency broadband (HFB) signal (see in refs. 8,38,39 for further details). We extracted visual-responsive contacts by a comparison of each contact's post-stimulus HFB response, averaged across the 100 to 500 ms period following stimulus onset, with its prestimulus baseline, averaged across the −400 ms to −100 ms interval prior to stimulus onset. This comparison was conducted using a two-tailed Wilcoxon signed-rank test, which was FDR corrected for all contacts (from all patients) together. Subsequently, contacts demonstrating a significant HFB response ($p_{fdr} < 0.05$) were categorized as visually responsive, totaling 377 contacts.

Following this, visually responsive contacts were grouped into subsets based on their anatomical and functional characteristics, namely face-selective ventral and fronto-parietal visual contacts. Face-selective ventral contacts were defined by comparing mean HFB responses between faces and places during the 100–500 ms post-stimulus window using a Wilcoxon rank-sum test. Contacts showing significantly greater activation to faces compared to places ($p_{fdr} < 0.05$), located anatomically beyond early visual regions (but excluding the frontal cortex), were categorized as face-selective ventral contacts ($n = 43$, depicted in blue in Fig. 1B). We defined the fronto-parietal ROI based on the Desikan Killiany atlas locations of the following labels: superior frontal gyrus, rostral middle frontal gyrus, pars orbitalis, pars triangularis, pars opercularis, precentral gyrus, postcentral gyrus, supramarginal gyrus, orbital frontal gyrus, and the anterior cingulate. Visual contacts located within these regions were categorized as the fronto-parietal visual contacts ($n = 63$, depicted in yellow in Fig. 1B). Any visually responsive contacts not falling into these specified categories were termed "other visually-responsive contacts" and are depicted in white in Fig. 1B, while non-visual contacts are depicted in gray. Further details can be found in ref. 8. A third group of visual contacts was extracted to be used as a second control for the Fronto-parietal ROI. We refer to this group as content-selective contacts ($n = 114$). Their extraction and results are fully described in the Supplementary Material and Supplementary Fig. 3.

Noise-celling measure. After dividing the visual contacts into different ROIs, we calculated a noise ceiling for each ROI. As ROIs were defined by pooling electrodes across participants, we used a split-half procedure based on stimulus repetitions to compute the noise ceiling. For each ROI and stimulus, we first calculated the average response of each contacts within the 0.1–0.4 s window (100–400 ms) following stimulus presentation. Next, we split the data into two halves, with two repetitions in each half. For each stimulus, we averaged the contact responses within each half. We defined the contacts' mean response population vectors as the stimulus activation patterns in each ROI. We then computed the Pearson correlation between the activation patterns of each pair of stimuli, resulting in an RDM. To estimate the reliability of the RDM, we calculated the correlation between the two halves. This process was repeated for all three possible split-half combinations, and the average correlation between the two halves was computed. The final average correlations were 0.77 for face-selective ventral contacts and 0.30 for fronto-parietal visual contacts.

Deep neural networks

We chose VGG, CLIP, and GPT-2 over newer models because we aim to balance state-of-the-art performance with architectures that are either purely visual, purely semantic, or multi-modal:

Visual network—ImageNet pre-trained VGG-16: To examine a network-based visual representation of the presented stimuli, we used the VGG-16 network[79] pre-trained on ImageNet[80], which includes 300,000 images from 1000 object categories. VGG is a commonly used network performing object and face categorization at human level[34,81–83]. We then extracted the embeddings of each image based on the feature vector representation in the penultimate- i.e., the hidden layer just below the output (fc7) layer of the network, which is the representation that is used for the classification. The similarity between each image pair was computed based on the cosine distance between these feature vectors.

Text network—openAI GPT2: A transformer-based language algorithm[84]. GPT is trained on a vast corpus of text to carry out various language tasks (e.g., semantic similarity, questions answering, grammatical correction, etc.), often involving longer text captions. Therefore, we chose to extract its representations using text from Wikipedia to obtain the representation of each stimulus (see Supplementary Table 1 for the text used for each stimulus). Specifically, we retrieved the embeddings of the first paragraph from Wikipedia corresponding to each familiar stimuli (identity or a place), based on the model output layer. Subsequently, we computed the similarity between each stimuli pair based on the cosine distance between the representations of these textual descriptions.

Multi-modal network - CLIP (Contrastive Language-Image pre-training)[85]: CLIP is trained in a self-supervised, contrastive manner to create similar representations for images and their text caption based on a training set of 400 M images and their captions from the internet[85]. We extracted the embeddings of each image based on the output layer of trained ResNet architecture using the visual and language components of CLIP.

CLIP Visual Component (CLIP-Image): We extracted the representations of the 28 visual images that were presented to the patients. We then computed the similarity between each object pair based on the cosine distance between these visual embeddings. We will refer to these representations as CLIP-Image.

CLIP Text component (CLIP-Text): CLIP textual component is trained with short captions that are used as labels of images on the internet and is limited to short text input (up to 77 characters). Therefore, we extracted its representations based on the names of the stimuli (see Supplementary Table 1 for the name used for each stimulus). We extracted the representations based on the names of the familiar exemplar (identity or a place) presented in each stimulus (e.g., "Eiffel Tower"). We computed the similarity between each object pair based on the cosine distance between these name representations. We will refer to these representations as CLIP-Text.

This provided us with the RDM of each network that is based on its output (or penultimate) layer. In addition, we also extracted the RDMs of each layer of the network.

As described above, we extracted the representations of GPT and CLIP-Text based on different text inputs. The representations of GPT were based on Wikipedia definitions, while the representations of CLIP-Text were based on the image's names. While we chose each input based on the network's nature of learning, extracting the representations of CLIP-Text based on names might seem like a significant reduction in information compared with GPT's input. However, we believe this is not the case. First, we extracted the similarity scores between all images and all names of the identities/places presented in the images using Image and CLIP-Text, respectively. This enabled us to create a familiarity RDM and identify which name was most similar to each image based on CLIP representation (see Supplementary Fig. 10). We found that for each image, the most similar name was its corresponding name (e.g., Barack Obama's image was closest to the text "Barack Obama" compared to any other name in the list), suggesting that CLIP's name-based representations carry relevant semantic information. Moreover, as seen in the results section, these CLIP-Text representations had the highest association with the fronto-parietal contacts, indicating that they represent the stimuli in some informative way. Finally, to test if we could extract further information from CLIP-Text, we also extracted its representations based on the Wikipedia definition (limited to 77 characters), which did not improve its association with the brain ROIs (see Supplementary Material for details).

Shuffled text representations. To assess CLIP-Text correlation with fronto-parietal contacts, we compared the original text representations with six variations of the same text, randomly shuffled. For each variation, we generated the representations based on CLIP-Text different layers and calculated the correlations with the fronto-parietal lobe in the first time window (0.1–0.4 s). Then we calculated the average of these correlations for each layer. Figure 5B shows the average correlation of the fronto-parietal contacts with CLIP-Text layers, based on these six shuffles.

Behavioral data: human similarity ratings

Participants. Forty participants were recruited for this study via the Prolific platform, with 20 participants assigned to each of the two experimental conditions: visual similarity based on images and semantic similarity based on names. The sample had a mean age of 29 years (SD = 2.9), with 20 females and one participant preferring not to disclose their gender. Participants received £6 for their participation (£9 per hour). Informed consent was obtained before the experiment, which was approved by the Tel Aviv University ethics committee, and all ethical regulations relevant to human research participants were followed. The sample size was determined based on prior research on similarity tasks[34].

Procedure. Participants rated the visual or semantic similarity of all 378 possible pairs of the 28 items:

Visual similarity task. On each trial, participants were shown a pair of images of two different items and asked to rate their visual similarity on a 6-point scale (1 = very similar, 6 = very dissimilar). The question, "How visually similar are these images?" appeared above the images, along with the similarity scale. Participants selected their ratings using a mouse. Each pair was displayed until the response, with the next pair appearing 1 s later. After every 80 pairs, participants took a mandatory 10-s break (four breaks total). After completing all 378 ratings, participants were asked to indicate their familiarity with each image before the experiment. The task took ~40 minutes to complete.

Semantic similarity task. Participants rated the similarity of item names based on their conceptual or semantic knowledge of the stimuli. The procedure and timing were identical to the visual task, except that the instruction was shown only at the beginning, without a trial-by-trial prompt.

Trial exclusion. Trials were excluded if participants indicated unfamiliarity with any item in the pair. Participants who were unfamiliar with 30% or more of the items were removed from the analysis entirely. Additionally, trials with response times shorter than 200 ms or longer than 30,000 ms (30 s) were excluded, as these were assumed to reflect poor task performance. In the visual similarity task, 23% of trials were excluded, including 33 due to response times exceeding 30 s, with the remainder excluded based on the familiarity criterion. In the semantic similarity task, 15% of trials were excluded, with 50 trials removed for exceeding the 30-s response time and the rest excluded due to the familiarity criterion.

Statistics and reproducibility (Data Analysis)

For the purpose of Representational Similarity Analysis (RSA) we created a dissimilarity matrix (RDM) of all possible stimuli pairs within the pool of the 28 stimuli (a total of 378 pairs), separately for each different brain ROI, the different DNNs and for the human similarity ratings.

Correlation between the DNN representations and brain ROIs. We first computed the Pearson correlation between the RDMs of each DNN and the RDMs of the ventral and fronto-parietal visual contacts. Then, to assess the unique contribution of each DNN to the neural responses, we computed the partial correlation of each DNN while holding the other DNNs constant. We calculated the partial correlation between each DNN and the brain data while controlling for all other DNNs, based on the general partial correlation formula. Since our study includes more than two DNNs, we used pairwise partial correlation, which measures the correlation between residuals—what remains in each variable after accounting for shared variance with others. For this, we applied the *pcor* function from the *ppcor* library[86] in R[87], for further details and equations of this process please see in ref. 86. Applying partial correlations provides us with the unique association between each pair of variables, that is not mediated by other variables, and this process is not affected by the order of the variables[88].

To test the significance of a correlation/partial correlation of a single network in a specific ROI we used a permutation test in which the networks' similarity scores (between pairs of stimuli) were held constant, but we shuffled the similarity scores that are brain-based. For each analysis and each ROI, we shuffled the similarity between pairs of stimuli in 10,000 iterations. Then we calculated the correlation/partial correlation (respectively) of the network with the patients' RDM (based on the shuffled similarities). P-value assigned to each network and layer was the proportion of iterations in which the original correlations/partial correlation (based on the real similarities) was greater (or smaller) than the value that was calculated in each iteration. *P* values were adjusted according to Phipson and Smyth[89] correction for permutation tests, and then they were FDR corrected for all networks that were tested in the same analysis, for each ROI separately. Then we used a two-tailed alpha (p value < 0.025) to infer significance. We also tested the significance of the difference between each pair of two DNNs or two time windows; To do so, we used a permutation test in which the networks' similarity scores were held constant, but we shuffled the similarity scores that are brain-based. For each analysis and each ROI, we shuffled the similarity between pairs of stimuli in 10,000 iterations. Then we calculated the correlation/partial correlation (respectively) of each network/time window with the neural RDM (based on the shuffled similarities), and proceeded to calculate the difference between each two networks'/time windows' correlations/partial correlations (with the neural RDMs). P value assigned to each comparison of a pair of networks was the proportion of iterations in which the original difference in correlations/partial correlation (based on the real similarities) was greater (or smaller) than the value that was calculated in each iteration. *P* values were adjusted according to[89] correction for permutation tests, and then they were FDR corrected for all networks that were tested in the same analysis, for each ROI separately. Then we used a two-tailed alpha to infer significance.

Correlation between the human similarity ratings and brain ROIs. Similarly to the DNNs, we now used the visual and semantic averaged RDM (across participants) as predictors of the ventral and fronto-parietal visual contacts response. We computed the Pearson correlation and partial correlations between the ROIs and the averaged RDM of each similarity task (when the second task is held out). To test the significance of a correlation/partial correlation of a single task in a specific ROI we used a permutation test in which the task similarity scores (between pair of stimuli) were held constant, but we shuffled the similarity scores that are brain-based. For each analysis and each ROI, we shuffled the similarity between pairs of stimuli in 10,000 iterations. Then we calculated the correlation/partial correlation (respectively) of the task with the patients' RDM (based on the shuffled similarities). P-value assigned to each task and was the proportion of iterations in which the original correlations/partial correlation (based on the real similarities) was greater (or smaller) than the value that was calculated in each iteration. P values were adjusted according to Phipson and Smyth[89] correction for permutation tests, and then they were FDR corrected. Then we used a two-tailed alpha to infer significance.

Data availability

The iEEG data and code underlying all main figures are available in this OSF folder[90]: https://doi.org/10.17605/OSF.IO/A2BJG. The original data[91] from Broday-Dvir et al. [8] is located in this Zenodo folder: https://doi.org/10.5281/zenodo.7813404.

Code availability

The code for extracting representations from DNNs is available upon request. The code for the data analysis, of the iEEG data after preprocessing and the behavioral data can be found in this OSF folder[90]: https://doi.org/10.17605/OSF.IO/A2BJG

Received: 17 March 2025; Accepted: 8 July 2025;
Published online: 30 July 2025

References

1. Grill-Spector, K. & Malach, R. The human visual cortex. *Annu. Rev. Neurosci.* **27**, 649–677 (2004).
2. Noy, N. et al. Ignition's glow: ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. *Conscious. Cogn.* **35**, 206–224 (2015).
3. Golland, Y. et al. Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. *Cereb. Cortex* **17**, 766–777 (2007).
4. Kronemer, S. I. et al. Human visual consciousness involves large scale cortical and subcortical networks independent of task report and eye movement activity. *Nat. Commun.* **13**, 7342 (2022).
5. Frässle, S., Sommer, J., Jansen, A., Naber, M. & Einhäuser, W. Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J. Neurosci.* **34**, 1738–1747 (2014).
6. Bellet, J. et al. Decoding rapidly presented visual stimuli from prefrontal ensembles without report nor post-perceptual processing. *Neurosci. Conscious.* **2022**, niac005 (2022).
7. Kapoor, V. et al. Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports. *Nat. Commun.* **13**, 1535 (2022).
8. Broday-Dvir, R., Norman, Y., Harel, M., Mehta, A. D. & Malach, R. Perceptual stability reflected in neuronal pattern similarities in human visual cortex. *Cell Rep*. **42**, 112614 (2023).
9. Vishne, G., Gerber, E. M., Knight, R. T. & Deouell, L. Y. Distinct ventral stream and prefrontal cortex representational dynamics during sustained conscious visual perception. *Cell Rep*. **42**, 112752 (2023).
10. Consortium, C. et al. Adversarial testing of global neuronal workspace and integrated information theories of consciousness. *Nature* **642**, 1–10 (2025).
11. Chan, A. W.-Y. & Downing, P. E. Faces and eyes in human lateral prefrontal cortex. *Front. Hum. Neurosci.* **5**, 51 (2011).
12. Machielsen, W. C. M., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P. & Witter, M. P. fMRI of visual encoding: reproducibility of activation. *Hum. Brain Mapp.* **9**, 156–164 (2000).
13. McDermott, K. B., Jones, T. C., Petersen, S. E., Lageman, S. K. & Roediger, H. L. Retrieval success is accompanied by enhanced activation in anterior prefrontal cortex during recognition memory: an event-related fMRI study. *J. Cogn. Neurosci.* **12**, 965–976 (2000).
14. Prendergast, G. et al. Differential patterns of prefrontal MEG activation during verbal & visual encoding and retrieval. *PLoS ONE* **8**, e82936 (2013).
15. Simony, E., Grossman, S. & Malach, R. Brain-machine convergent evolution: a window into the functional role of neuronal selectivity. *Proc. Natl. Acad. Sci. USA* **121**, e2319709121 (2023).
16. Kanwisher, N., Khosla, M. & Dobs, K. Using artificial neural networks to ask 'why'questions of minds and brains. *Trends Neurosci.* **46**, 240–254 (2023).
17. Bomatter, P. et al. When pigs fly: Contextual reasoning in synthetic and natural scenes. In *Proc. IEEE/CVF International Conference on Computer Vision* 255–264 (IEEE, 2021).
18. Vinken, K., Boix, X. & Kreiman, G. Incorporating intrinsic suppression in deep neural networks captures dynamics of adaptation in neurophysiology and perception. *Sci. Adv.* **6**, eabd4205 (2020).
19. Ponce, C. R. et al. Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* **177**, 999–1009 (2019).
20. Ritchie, J. B. et al. Untangling the animacy organization of occipitotemporal cortex. *J. Neurosci.* **41**, 7103–7119 (2021).
21. Wardle, S. G. & Baker, C. I. Recent advances in understanding object recognition in the human brain: deep neural networks, temporal dynamics, and context. *F1000Research* **9**, F1000 (2020).
22. Bankson, B. B., Hebart, M. N., Groen, I. I. A. & Baker, C. I. The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks. *Neuroimage* **178**, 172–182 (2018).
23. Kriegeskorte, N. *Deep Neural Networks: A New Framework for Modelling Biological Vision and Brain Information Processing*. http://biorxiv.org/lookup/doi/10.1101/029876 (2015).
24. Malach, R. Local neuronal relational structures underlying the contents of human conscious experience. *Neurosci. Conscious.* **2021**, niab028 (2021).
25. Edelman, S., Grill-Spector, K., Kushnir, T. & Malach, R. Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* **26**, 309–321 (1998).
26. Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis-connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 249 (2008).
27. Lau, H., Michel, M., LeDoux, J. E. & Fleming, S. M. The mnemonic basis of subjective experience. *Nat. Rev. Psychol.* **1**, 479–488 (2022).
28. Grossman, S. et al. Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nat. Commun.* **10**, 1–13 (2019).
29. Dobs, K., Martinez, J., Kell, A. J. E. & Kanwisher, N. Brain-like functional specialization emerges spontaneously in deep neural networks. *Sci. Adv.* **8**, eabl8913 (2022).
30. Groen, I. I. A. Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *Elife* **7**, e32962 (2018).
31. Hasson, U., Nastase, S. A. & Goldstein, A. Direct fit to nature: an evolutionary perspective on biological and artificial neural networks. *Neuron* **105**, 416–434 (2020).
32. Carlin, J. D. & Kriegeskorte, N. Adjudicating between face-coding models with individual-face fMRI responses. *PLoS Comput Biol.* **13**, 1–28 (2017).
33. Schyns, P. G., Snoek, L. & Daube, C. Degrees of algorithmic equivalence between the brain and its DNN models. *Trends Cogn. Sci.* **26**, 1090–1102 (2022).
34. Shoham, A., Grosbard, I. D., Patashnik, O., Cohen-Or, D. & Yovel, G. Using deep neural networks to disentangle visual and semantic information in human perception and memory. *Nat. Hum. Behav*. **8**, 702–717 (2024).
35. Wang, A. Y., Kay, K., Naselaris, T., Tarr, M. J. & Wehbe, L. Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset. *Nat. Mach. Intell.* **5**, 1415–1426 (2023).
36. Doerig, A. et al. Semantic scene descriptions as an objective of human vision. *ArXiv Prepr. ArXiv220911737* https://arxiv.org/abs/2209.11737 (2022).
37. Subramaniam, V. et al. Revealing Vision-Language Integration in the Brain with Multimodal Networks. Preprint at http://arxiv.org/abs/2406.14481 (2024).

38. Norman, Y. et al. Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. *Science* **365**, eaax1030 (2019).
39. Norman, Y., Yeagle, E. M., Harel, M., Mehta, A. D. & Malach, R. Neuronal baseline shifts underlying boundary setting during free recall. *Nat. Commun.* **8**, 1301 (2017).
40. Consortium, C. et al. An adversarial collaboration to critically evaluate theories of consciousness. *BioRxiv* 2023–06 https://doi.org/10.1101/2023.06.23.546249 (2023).
41. Goldberg, I. I., Harel, M. & Malach, R. When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* **50**, 329–339 (2006).
42. Hesselmann, G., Hebart, M. & Malach, R. Differential BOLD activity associated with subjective and objective reports during "blindsight" in normal observers. *J. Neurosci.* **31**, 12936–12944 (2011).
43. Malach, R. The role of the prefrontal cortex in conscious perception: the localist perspective. *J. Conscious. Stud.* **29**, 93–114 (2022).
44. Tsuchiya, N., Wilke, M., Frässle, S. & Lamme, V. A. No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* **19**, 757–770 (2015).
45. Stuss, D. T. & Benson, D. F. Neuropsychological studies of the frontal lobes. *Psychol. Bull.* **95**, 3 (1984).
46. Corbetta, M., Kincade, M. J., Lewis, C., Snyder, A. Z. & Sapir, A. Neural basis and recovery of spatial attention deficits in spatial neglect. *Nat. Neurosci.* **8**, 1603–1610 (2005).
47. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* https://doi.org/10.1038/nn.4244. (2016)
48. Cohen, U., Chung, S., Lee, D. D. & Sompolinsky, H. Separability and geometry of object manifolds in deep neural networks. *Nat. Commun.* **11**, 746 (2020).
49. Doerig, A. et al. The neuroconnectionist research programme. *Nat. Rev. Neurosci.* **24**, 431–450 (2023).
50. Panagiotaropoulos, T. I. An integrative view of the role of prefrontal cortex in consciousness. *Neuron* **112**, 1626–1641 (2024).
51. Quiroga, R. Q. Concept cells: the building blocks of declarative memory functions. *Nat. Rev. Neurosci.* **13**, 587–597 (2012).
52. Quiroga, R. Q., Mukamel, R., Isham, E. A., Malach, R. & Fried, I. Human single-neuron responses at the threshold of conscious recognition. *Proc. Natl Acad. Sci. USA* **105**, 3599–3604 (2008).
53. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303**, 1634–1640 (2004).
54. Pitcher, D., Garrido, L., Walsh, V. & Duchaine, B. C. Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *J. Neurosci.* **28**, 8929–8933 (2008).
55. Berkovich-Ohana, A. et al. Inter-participant consistency of language-processing networks during abstract thoughts. *NeuroImage* **211**, 116626 (2020).
56. Shoham, A., Broday-Dvir, R., Malach, R. & Yovel, G. The organization of high-level visual cortex is aligned with visual rather than abstract linguistic information. *bioRxiv* 2024–11 https://doi.org/10.1101/2024.11.12.623145 (2024).
57. Grill-Spector, K. et al. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* **24**, 187–203 (1999).
58. Grill-Spector, K., Henson, R. & Martin, A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10**, 14–23 (2006).
59. Gerber, E. M., Golan, T., Knight, R. T. & Deouell, L. Y. Cortical representation of persistent visual stimuli. *Neuroimage* **161**, 67–79 (2017).
60. Golan, T. et al. Human intracranial recordings link suppressed transients rather than'filling-in'to perceptual continuity across blinks. *elife* **5**, e17243 (2016).
61. Melloni, L., Mudrik, L., Pitts, M. & Koch, C. Making the hard problem of consciousness easier. *Science* **372**, 911–912 (2021).
62. Lamme, V. A. Visual functions generating conscious seeing. *Front. Psychol.* **11**, 83 (2020).
63. Lamme, V. A. & Roelfsema, P. R. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* **23**, 571–579 (2000).
64. Mashour, G. A., Roelfsema, P., Changeux, J.-P. & Dehaene, S. Conscious processing and the global neuronal workspace hypothesis. *Neuron* **105**, 776–798 (2020).
65. Dehaene, S. & Changeux, J.-P. Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200–227 (2011).
66. Brown, R., Lau, H. & LeDoux, J. E. Understanding the higher-order approach to consciousness. *Trends Cogn. Sci.* **23**, 754–768 (2019).
67. Lau, H. & Rosenthal, D. Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* **15**, 365–373 (2011).
68. Michel, M. Conscious perception and the prefrontal cortex a review. *J. Conscious. Stud.* **29**, 115–157 (2022).
69. Block, N. What is wrong with the no-report paradigm and how to fix it. *Trends Cogn. Sci.* **23**, 1003–1013 (2019).
70. Wang, X., Men, W., Gao, J., Caramazza, A. & Bi, Y. Two forms of knowledge representations in the human brain. *Neuron* **107**, 383–393.e5 (2020).
71. Riberto, M., Paz, R., Pobric, G. & Talmi, D. The neural representations of emotional experiences are more similar than those of neutral experiences. *J. Neurosci.* **42**, 2772–2785 (2022).
72. Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S.-M. & Carlson, T. A. Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *NeuroImage* **132**, 59–70 (2016).
73. Tsantani, M. et al. FFA and OFA encode distinct types of face identity information. *J. Neurosci.* **41**, 1952–1969 (2021).
74. Bainbridge, W. A., Hall, E. H. & Baker, C. I. Distinct representational structure and localization for visual encoding and recall during visual imagery. *Cereb. Cortex* **31**, 1898–1913 (2021).
75. Naspi, L., Hoffman, P., Devereux, B. & Morcom, A. M. Perceptual and semantic representations at encoding contribute to true and false recognition of objects. *J. Neurosci.* **41**, 8375–8389 (2021).
76. Bayer, M., Berhe, O., Dziobek, I. & Johnstone, T. Rapid neural representations of personally relevant faces. *Cereb. Cortex* **31**, 4699–4708 (2021).
77. Li, Y., Zhang, M., Liu, S. & Luo, W. EEG decoding of multidimensional information from emotional faces. *NeuroImage* **258**, 119374 (2022).
78. Katabi, N. et al. Deeper than you think: partisanship-dependent brain responses in early sensory and motor brain regions. *J. Neurosci.* **43**, 1027–1037 (2023).
79. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. https://doi.org/10.48550/arXiv.1409.1556 (2014)
80. Deng, J. ImageNet: A large-scale hierarchical image database. In *Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition 248–255* (IEEE, https://doi.org/10.1109/CVPR.2009.5206848. 2009).
81. Jacob, G., Pramod, R. T., Katti, H. & Arun, S. P. Qualitative similarities and differences in visual object representations between brains and deep networks. *Nat. Commun.* **12**, 1872 (2021).
82. Abudarham, N., Grosbard, I. & Yovel, G. Face recognition depends on specialized mechanisms tuned to view-invariant facial features: insights from deep neural networks optimized for face or object recognition. *Cogn. Sci.* **45**, 13031 (2021).
83. Yovel, G., Grosbard, I. & Abudarham, N. Deep learning models challenge the prevailing assumption that face-like effects for objects of expertise support domain-general mechanisms. *Proc. R. Soc. B* **290**, 20230093 (2023).

84. Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. Improving language understanding by generative pre-training.
85. Radford, A. Learning transferable visual models from natural language supervision. In *Proc. 38th International Conference on Machine Learning* (eds. Meila, M. & Zhang, T.) 8748–8763 (PMLR, 2021).
86. Kim, S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **22**, 665 (2015).
87. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2024).
88. Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Routledge, 2013).
89. Phipson, B. & Smyth, G. K. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol*. **9**, Article39 (2010).
90. Shoham, A. et al. Data and code supporting the paper: Text-related functionality and dynamics of visual human pre-frontal activations revealed through neural network convergence. https://doi.org/10.17605/OSF.IO/A2BJG (2025).
91. Broday-Dvir, R. Norman, Y. Harel, M. Mehta, A., & Malach, R. Data and code related to the manuscript 'Perceptual stability reflected in neuronal pattern similarities in human visual cortex'. Zenodo https://doi.org/10.5281/zenodo.7813404 (2023).

Acknowledgements

The study was funded by a CIFAR BMC fellowship to R. Malach, ISF grant no.917/21 to G. Yovel, and by the Sieratzki Institute for Advances in Neuroscience fellowship to A.S.

Author contributions

Conceptualization—A.S., R.B.D., I.Y., R.M., and G.Y.; Methodology—A.S., R.M., and G.Y.; Software—A.S., and R.B.D.; Formal Analysis—A.S., and R.B.D.; Data Curation—A.S. and R.B.D.; Writing—Original Draft—A.S. and R.M.; Writing—Review and Editing—A.S., R.B.D., I.Y., R.M., and G.Y.; Visualization—A.S.; Supervision—R.M. and G.Y.; Project Administration—R.M. and G.Y.; Funding Acquisition—R.M., G.Y., and A.S.

Competing interests

The author declares no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-08497-8.

Correspondence and requests for materials should be addressed to Adva Shoham or Rafael Malach.

Peer review information *Communications Biology* thanks Yaara Erez and Joseph Neisser for their contribution to the peer review of this work. Primary Handling Editors: Helen Blank and Jasmine Pan. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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