Systems/Circuits

Face-Selective Units in Human Ventral Temporal Cortex Reactivate during Free Recall

Simon Khuvis,[1] Erin M. Yeagle,[1] Yitzhak Norman,[2] Shany Grossman,[2] Rafael Malach,[2] and Ashesh D. Mehta[1]

[1]Department of Neurosurgery, Feinstein Institutes for Medical Research, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Manhasset, New York 11030, and [2]Department of Neurobiology, Weizmann Institute of Science, Rehovot, 76100, Israel

Research in functional neuroimaging has suggested that category-selective regions of visual cortex, including the ventral temporal cortex (VTC), can be reactivated endogenously through imagery and recall. Face representation in the monkey face-patch system has been well studied and is an attractive domain in which to explore these processes in humans. The VTCs of 8 human subjects (4 female) undergoing invasive monitoring for epilepsy surgery were implanted with microelectrodes. Most (26 of 33) category-selective units showed specificity for face stimuli. Different face exemplars evoked consistent and discriminable responses in the population of units sampled. During free recall, face-selective units preferentially reactivated in the absence of visual stimulation during a 2 s window preceding face recall events. Furthermore, we show that in at least 1 subject, the identity of the recalled face could be predicted by comparing activity preceding recall events to activity evoked by visual stimulation. We show that face-selective units in the human VTC are reactivated endogenously, and present initial evidence that consistent representations of individual face exemplars are specifically reactivated in this manner.

Key words: face; free recall; fusiform face area; imagery; single unit; vision

Significance Statement

The role of "top-down" endogenous reactivation of native representations in higher sensory areas is poorly understood in humans. We conducted the first detailed single-unit survey of ventral temporal cortex (VTC) in human subjects, showing that, similarly to nonhuman primates, humans encode different faces using different rate codes. Then, we demonstrated that, when subjects recalled and imagined a given face, VTC neurons reactivated with the same rate codes as when subjects initially viewed that face. This suggests that the VTC units not only carry durable representations of faces, but that those representations can be endogenously reactivated via "top-down" mechanisms.

Introduction

Facial recognition is an essential adaptive social function in primates, facilitated by the extensive development of specialized visual areas in the brain's ventral temporal cortex (VTC). Information processing in this region must meet social demands to perceive, classify and uniquely identify a multitude of faces. First described in monkeys (Gross et al., 1969, 1972; Desimone et al., 1997), studies of single neurons firing in response to visual features of faces have uncovered key "bottom-up" mechanisms of the feature space that drive neuronal activity (Leopold et al., 2006; Tsao et al., 2006, 2008; DiCarlo et al., 2012; Yamins et al., 2014; Afraz et al., 2015; Chang and Tsao, 2017). VTC neurons exhibit complex facial feature sensitivity, supporting their role in the discrimination of individual face exemplars. However, their role in nonsensory, extraretinal processing, as would occur during imagery and recall, is difficult to determine from monkeys, who cannot communicate their subjective experiences in detail. In humans, neuroimaging has defined a critical node in face processing within the VTC: the fusiform face area (FFA) (Kanwisher et al., 1997). fMRI studies support both a sensory

Received Dec. 1, 2019; revised Aug. 15, 2020; accepted Sep. 8, 2020.

Author contributions: S.K., Y.N., and R.M. designed research; S.K., E.M.Y., and A.D.M. performed research; S.K., E.M.Y., and S.G. analyzed data; S.K. and A.D.M. wrote the first draft of the paper; S.K., E.M.Y., Y.N., S.G., R.M., and A.D.M. edited the paper.

A.D.M. was supported by National Institutes of Health/National Institute of Neurological Disorders and Stroke NS098976-01 and NSF-BSF-2017015; and National Institute of Mental Health MH114166-01. S.K. was supported by the Feinstein M.D. Scholarship. This work was supported in part by US-Israel BSF Grant to A.D.M. and R.M. We thank Salman Qasim and Drs. Joshua Jacobs, Ida Momennejad, Elliot Smith, Ella Podvalny, Tal Golan, Pierre Megévand, José Herrero, Victor Du, and Stephan Bickel for intellectual and technical support; Dr. Xiangzhi Zhou for help with the fMRI experiments; our colleagues at the Northwell Health Comprehensive Epilepsy Center, especially Willie Walker, Monika Lalik, and Drs. Sean Hwang, Fred Lado, and Scott Stevens; and our patient-subjects and their families, without whose patience and cooperation, this research would not have been possible.

S. Khuvis' present address: Department of Psychiatry, New York University Grossman School of Medicine and Center for Neural Science, New York University, 550 First Avenue, New York, NY 10016.

A.D.M. has a consulting agreement with P.M.T. Corporation. The remaining authors declare no competing financial interests.

Correspondence should be addressed to Ashesh D. Mehta at amehta@northwell.edu.

https://doi.org/10.1523/JNEUROSCI.2918-19.2020

Table 1. Demographic and recording information for all participating subjects[a]

									1-back						Free recall		
Subject	Sex	Age (yr)	Handedness	Primary language	FSIQ	Faces	Implant no.	Lat	Vis	F	B	T	H	P	Vis	F	S
1	F	32	R	English	67	4	1	L	2	2	0	0	0	0			
2	M	41	L	Spanish	N/A	N/A	1	R	7	0	0	0	0	0	14	0	2
							2	L	4	1	1	0	0	0	10	5	0
3	F	55	R	English	46	NT	1	R	8	8	0	0	0	0	12	7	0
								L	1	0	0	0	0	0	2	0	0
4	F	37	R	English	74	2	1	L	5	0	0	1	0	0			
5	M	20	R	Spanish	N/A	N/A	1	R	8	0	0	0	2	1			
6	F	32	R	English	88	10	1	R	13	3	0	0	2	0	8	0	1
7	M	43	R	English	91	12	1	R	2	0	0	0	0	0			
								L	1	0	0	0	0	0			
8	M	20	R	English	77	11	1	R	12	12	0	0	0	0	16	16	0

[a]FSIQ, full-scale IQ, as measured by the Wechsler Adult Intelligence Scale, Ed 4; Lat, laterality of electrode; Vis, visually-responsive units; F, face-selective units; B, body part-selective units; T, tool-selective units; H, house-selective units; P, pattern-selective units (the sole pattern selective unit shows a weaker response to patterns than all other stimulus classes); S, scene- (place) selective units; NT, not tested. Pearson Clinical (London), Faces: Face (supplemental) scaled score from "Pearson Clinical"; "Social Cognition Test", a test that requires intact facial recognition, memory, and concentration (mean of 10, SD of 3).

("bottom-up") as well as a cognitive ("top-down") role for the FFA, which is activated not only when subjects view faces, but also when they expect to see a face (Puri et al., 2009; Bollinger et al., 2010), perform imagery tasks involving faces (O'Craven and Kanwisher, 2000; Ishai et al., 2002), and hold face representations in working memory (Ranganath et al., 2004). Activation of category-selective regions of VTC can predict recall of items in that category (Polyn et al., 2005; Norman et al., 2017). In addition to fMRI studies, MEG (Liu et al., 2002) and direct electrocorticographic recordings (Singer et al., 2015; Jacques et al., 2016) in humans show fusiform responses emerging early (~100 ms) after face presentation, suggesting a sensory role for this region. While these studies have provided important insights, they lack the single-cell resolution needed to uncover how bottom-up, feature-based, sensory processing relates to top-down processes at the level of single human neurons. Thus, the complex neuronal selectivity for facial features revealed by monkey neurophysiology, the relatively early sensitivity revealed by human field potential recordings and the fMRI evidence for top-down control of the FFA have yet to be integrated.

Clinical macroelectrodes with microwires provide an opportunity to investigate this question by studying isolated neuronal spiking in patients undergoing chronic recordings. Such *in vivo* human data have provided insights into a number of physiological and pathologic processes, most notably in the medial temporal lobe (Quiroga et al., 2005). Face-responsive units have been reported in the human VTC (Axelrod et al., 2019), but their single and population spiking activity has yet to be explored in detail. In experiments described here, we recorded from microwires in the VTCs of human subjects as they viewed face stimuli and later recalled them in an episodic free recall task, allowing us to examine human sensory and higher-order cognitive processes at the single-unit level. Recorded units showed a diversity of response patterns while maintaining strong category specificity. We show that population responses to the presentation of different face exemplars can be robustly discriminated, and we present evidence that these responses are reinstated when subjects recall and visualize previously presented face images in the absence of sensory input. Our results support models of memory in which single-neuronal substrates of sensory processing are reactivated in a top-down fashion during recall (Cowan, 1988).

Materials and Methods

Software accessibility. Code and processed data are available on: https://github.com/IEEG/SUFreeRecall. Raw data will be provided on request.

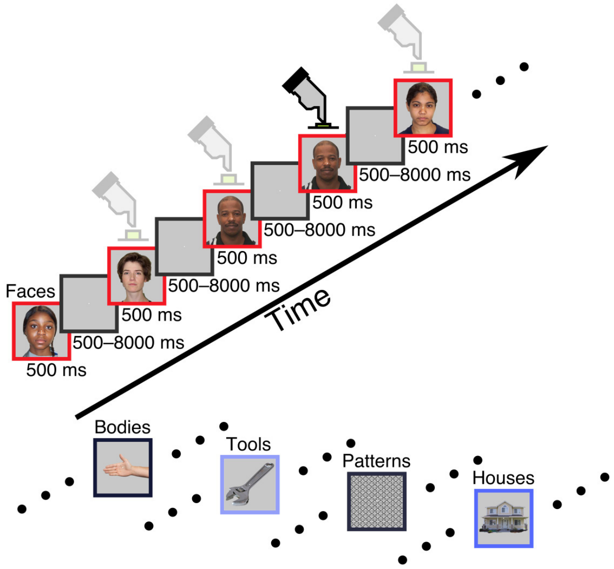

Figure 1. One-back task design. Subjects were shown a series of images of one of five categories: faces, bodies, houses, patterns, and tools. Each image was displayed for 500 ms. There was a random interstimulus interval of 500-8000 ms. Blocks of images of the same category were shown together. Subjects were instructed to press the button when they saw the same image twice in a row.

Subject recruitment. Study protocols were approved by the Institutional Review Board of the Feinstein Institutes for Medical Research. Eight subjects (4 females) scheduled to undergo intracranial EEG (iEEG) of the inferior occipitotemporal cortex, including either the right or left FFA, for diagnosis of focal epilepsy were recruited. The determination of which brain areas to monitor for each subject was made on clinical criteria. One subject (Subject 2) was implanted twice: first on the right and then on the left. He performed the 1-back task twice, and three runs of the free recall task (typically there are two runs per subject).

Six subjects had electrodes on the right, 5 subjects had electrodes on the left (three bilateral). Subject demographic and surgical data and recording yields are presented in Table 1.

Electrode placement. Subjects 3–8 performed the same 1-back task during an fMRI scan that they did during intracranial single-unit recordings (see Fig. 1; for task details, see Experimental design: tasks). Focal areas with the greatest Face>House contrast within the fusiform gyrus, and consistent with the expected anatomy, were identified as the FFA.

In the recruited subjects, monitoring of brain areas, including the FFA at least unilaterally, was clinically indicated for localization of epilepsy and mapping of functional areas. Precise electrode trajectories are typically defined by clinical circumstances; and when a range of a priori equally efficacious and safe alternatives included a trajectory leading to either the right or left FFA, then this brain area was targeted by (1) anatomic approximation based on preoperative structural MRI in the Radionics electrode trajectory planning software (Integra LifeSciences) by a neurosurgeon experienced in functional neuroanatomy (A.D.M.), based on landmarks known to define the FFA (Weiner et al., 2014); (2) at least two investigators (A.D.M. and S.K.) comparing the relative anatomy of the fMRI-defined FFA with the anatomy shown in the preoperative MRI; or (3) superimposing the Face>House contrast map from the fMRI over the preoperative MRI in the StealthStation trajectory planning software (Medtronic). The end of each macro electrode was planned to stop 4 mm short of the centroid of the targeted area so that the microwires, when deployed, would terminate in the ROI.

Ad-Tech Medical micro-macro depth electrodes (Misra et al., 2014) were used to simultaneously obtain clinical EEG data as well as record single-unit activity with minimal additional surgical risk. Micro-electrodes were trimmed to 4 mm beyond the end of the macro canula before insertion. One micro-macro electrode was implanted in each targeted FFA, each with eight signal and one reference microwires.

MRI colocalization. Six of the 8 subjects underwent an fMRI scan in addition to the preimplant anatomic image acquisition. BOLD contrast was obtained using a T2* sensitive EPI sequence: TR 2000 ms; TE 30 ms; flip angle 77°; FOV 192 mm; voxel size 2.1 × 2.1 × 2.1 mm; 34 slices). During the functional scan, subjects performed the same 1-back task as performed during iEEG recordings (see Fig. 1; Experimental design: tasks).

fMRI data analyses were conducted using FSL's FEAT (version 6.00; www.fmrib.ox.ac.uk/fsl) (Jenkinson et al., 2012). Preprocessing steps included discarding first two acquired volumes; motion correction using MCFLIRT, with the first volume as a reference; slice-timing correction of voxels time series using sinc interpolation; high-pass temporal filtering with a cutoff of 0.0083 Hz (maximal cycle of 120 s); and spatial smoothing of each acquired volume using a Gaussian kernel with FWHM of 2 mm.

A GLM was fit to each subject's run, with blocks from the five different visual categories modeled by five independent box car predictors convolved with a double γ HRF model. Motion correction time series, their first derivatives, and their squared time series were included as additional confound predictors (18 predictors in total, 6 original motion series ×3). Each predictor and voxel's time series were demeaned before computing the GLM to account for baseline activity level. For each condition and contrast, we obtained a whole-brain t value map, and aligned the map to the high-resolution anatomic image of the subject using FLIRT.

Electrodes were localized using the iELVis toolbox (Groppe et al., 2017). A postimplant CT for each subject was registered to the preimplant MRI via the postimplant MRI, using FSL's FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002; Greve and Fischl, 2009). Following coregistration, electrode artifacts were identified manually in BioImageSuite (Papademetris et al., 2006).

For visualization, functional data were coregistered to the subject's brainmask volume (generated via FreeSurfer's recon_all) using 3dAllineate (AFNI toolbox; www.afni.nimh.nih.gov) (Cox, 1996) and thresholded (minimum t value = 1.5–3; maximum t value = 6, cluster threshold = 100 voxels). The contrast used was faces versus houses, except for 1 subject (Subject 4) for whom faces versus patterns yielded a clearer map. Macroelectrode coordinates were superimposed on the contrast map using BioImageSuite (Papademetris et al., 2006).

Data acquisition. Neural data were acquired at 30 kHz on a Blackrock Microsystems amplifier from either one or both FFAs of each participating subject, using a Blackrock Microsystems Cabrio headstage to current-amplify and rereference each channel to the reference wire in its respective bundle. Recordings took place between 1 and 3 d after implant to minimize the possibility of microwire damage and unit dropout.

Experimental design: tasks. All subjects performed the 1-back task shown in Figure 1. Sets of 10 faces, body parts, houses, patterns, and tools were presented centrally, on a laptop monitor using Presentation software (version 0.70, Neurobehavioral Systems; picture size: ~12° at ~60 cm viewing distance) while subjects were sitting up in bed. Images were shown in blocks of a single category at a time. All 10 exemplars from the relevant category were presented once in each block in a pseudo-random order, with the exception of 1-back repeats where the same exemplar was presented twice. Each image was presented for 0.50 s, followed by a jittered interstimulus period of between 0.75 and 1.5 s. Blocks were separated by either 4 or 8 s. Two versions of the task were used. One contained 250 trials (25 blocks), and was only performed by Subject 4, and the other contained 260 trials (26 blocks), and was performed by all other subjects. Each exemplar was presented 3-8 times throughout the task. There were 18 or 19 1-back repetitions during the task, for which the subjects were instructed to click a mouse button on the laptop on which the images were presented. Faces images were drawn from the dataset compiled by Minear and Park (2004).

The following are adapted from Norman et al. (2017): 4 subjects (Subjects 2, 3, 6, and 8) performed the free recall task. The experiment was divided into two runs. Participants were presented with 14 different images of famous faces and popular landmarks. Each image repeated 4 times (1.5 s duration, 0.75 s interstimulus interval) in a pseudorandom order, such that each presentation cycle contained all of the different images, but the order of images was randomized within the cycle. The same image was never presented twice consecutively. Participants were instructed to look carefully at the images and try to remember them in detail, emphasizing unique colors, unusual face expressions, perspective, and so on. Stimuli were presented on a standard laptop liquid crystal display screen using the Presentation software (Version 0.70, Neurobehavioral Systems; picture size: 17° × 13° at ~60 cm viewing distance).

After viewing the pictures, participants put a blindfold on and began a short interference task of counting back from 150 in steps of 5 for 40 s. Upon completion, recall instructions were presented, guiding the subjects to recall items from only one category at a time, starting with faces in the first run and with places in the second run, and to verbally describe each image they recall, as soon as it comes to mind, with two or three prominent visual features. The instructions also emphasized reporting everything that came to mind during the free recall period. The duration of the free recall phase was 2.5 min per each category (5 min in total, × 2 runs). In case the subjects indicated that they could not remember any more items, they received a standard prompt from the experimenter (e.g., "Can you remember any more pictures?"). The order of the recalled categories was fixed across subjects and counterbalanced between the two runs. A different set of images (7 per category) was presented in each run.

Verbal responses during the free recall phase were continuously recorded via a microphone attached to the subject's gown. The onset and offset of each recall event were extracted in an offline analysis, identifying the first/last soundwave amplitude change relevant to each utterance using Audacity recording and editing software (version 2.0.6; https://www.audacityteam.org/).

Spike detection and sorting. Combinato (Niediek et al., 2016) was used for spike detection and sorting in each of the channels. Candidate units were visually inspected for waveform consistency and plausible interspike interval distribution. When ambiguous, we preferred not to overscreen at this stage and to allow for subsequent objective analyses to reject artifacts.

Units were visually reviewed, and those with robust waveform/interspike interval characteristics were labeled as single units, those with waveforms consistent with discrete action potentials (but with a wider variance) were labeled as multiunits. In this context, the origins of multiunit deflections may well be single cells, but because of noise, they are not demonstrably well isolated. The subsequent analyses are agnostic to single-unit or multiunit designations, and we will refer to them collectively as "units" unless a specific unit is being described.

Candidate units that did not have waveforms consistent with discrete action potentials were discarded from further analysis, and units with identical or extremely similar waveforms within a single channel were merged together.

We do not merge single units back into their respective multiunits once they have been separated, so we have taken measures to prevent the effects of potential overclustering from biasing our results as we describe in Materials and Methods.

Statistical analyses: identification of visually responsive units in the 1-back task. Spike trains were aligned to stimulus onset. The ON period was defined as 0.1–0.5 s after each stimulus appeared on screen and the OFF period was defined as 0.6-1 s after onset (0.1–0.5 s after the reappearance of the fixation cross). If at least one spike was not recorded during nine or more ON or OFF periods, then the unit was discarded from further analysis.

A Friedman's test was used to determine whether there was a significant effect of time relative to stimulus onset on unit firing rates. The test was performed on the *n* by *i* array: $|\Delta R_i[n]|$, where $R_i[n]$ is the firing rate in each bin index *n* of width 0.033 s from 0 to 0.5 s after stimulus change (onset or offset), Δ is the first discrete differential, and *i* is the run number, used as the multiple measures factor in the Friedman's test. The test is performed relative to onset and offset, and compared with $\alpha = 0.05$ with Bonferroni correction for multiple comparisons across units and both periods. It has the effect of looking for changes in time-varying firing rates while ignoring amplitude differences across stimuli. Units were classified as ON cells if they were visually responsive during the ON period, and OFF cells if and only if they were visually responsive during the OFF period and not the ON period. Because of the potential persistence of offset responses into the interstimulus interval, we did not use a prestimulus baseline; instead, the immediate poststimulus but preresponse period functions as a baseline, a method also used in human evoked response potential research (e.g., Mossbridge et al., 2013) and in monkey single-unit electrophysiology (Tsao et al., 2006). To avoid specifying a latency *a priori* and recognizing that the response times for different neurons may be different, we chose to compare the bins of the response raster to each other rather than to a specific baseline period. This carries the advantage of making the procedure more sensitive to neurons with brief, peaked responses, but may make it less sensitive to neurons with early and very uniformly sustained responses. A manual review of the data processed as described showed that this technique generally did not omit clearly visually responsive neurons (although one obvious OFF unit with a short latency and uniform firing rate was missed), and with a sufficient threshold for detection, did not include units that were not clearly responsive.

We also performed a supplementary analysis using a more conventional Wilcoxon rank-sum test comparing the 0.4 s ON period to the 0.4 s before stimulus onset for each category in each unit, combining across categories using Fisher's method of combined probabilities and correcting for multiple comparisons using the Bonferroni–Holm method. We compared the consistency between the classification of units (excluding offset-responsive units) as visually responsive using this method at the best α threshold and the method described previously.

Identification of category-selective units in the 1-back task. A unit was defined as category selective if spike rates were significantly unequal among stimulus categories (sample sizes for Subjects 1–3 and 5–8: 60 face presentation trials and 50 presentations trials of all other stimulus categories; sample sizes for Subject 4: 50 presentation trials of all stimulus categories) during the ON and OFF periods for ON and OFF cells, respectively, according to a Kruskal–Wallis one-way ANOVA test at $\alpha = 0.05$ corrected for multiple comparisons across visually responsive units using the Bonferroni–Holm technique. Firing rates among categories were further compared with establish if any one category showed significantly higher or lower firing rates than all others at $\alpha = 0.05$, using the Fisher's least significant difference procedure to account for multiple comparisons; if one category of stimuli evoked significantly stronger or weaker responses, then the unit was said to be selective for that category of stimuli, and "excited" or "inhibited," respectively. If a unit simultaneously showed significant increases and decreases to different categories, then the magnitudes of the responses (absolute values of the natural-log-ratios of the spikes rates during the ON period and the 0.4 s before stimulus onset for ON cells and during the ON and OFF periods for OFF cells, with 0.1 Hz pseudo-counts added before taking the log) are compared using a Wilcoxon rank-sum test with stimulus categories as groups. If responses to one category are significantly stronger than to the other at $\alpha = 0.05$ (two-tailed), then the unit is said to be selective for that category.

Calculation of unit response latency. Response latency of face-selective units was calculated by constructing peristimulus time histograms with bin widths of 0.02 s for each face-selective unit, and comparing the activity 0 and $0.9\overline{6}$ s after onset to Z^{-1} (0.95) × S_{BL} for ON cells, where Z^{-1} is the inverse normal cumulative distribution function, and S_{BL} is the sample SD of the firing rate among the bins between 0.5 and 0 s before stimulus onset. The center of the first of three consecutive bins exceeding this threshold is defined as the onset latency for that unit.

Relationship between units' visual responsiveness and category selectivity. The χ^2 parameter from the Friedman test used to measure unit responsiveness (responsiveness index) is plotted against the χ^2 parameter from the Kruskal–Wallis test used to measure unit category selectivity (selectivity index) on a log-log scale. We used only subjects with the same distribution of trials per exemplar so that category selectivity values could be directly compared. The Spearman correlation between the log values of the two indices is calculated and tested against zero at $\alpha = 0.05$.

The procedure is repeated using the d' face sensitivity index instead of the χ^2 Kruskal–Wallis value.

Exemplar decoding. Log-firing rates (pseudo-count 1 added before taking the log) from pseudo-populations of visually responsive units from all 8 subjects, at 0.1–0.5 s after image presentation or offset (for ON and OFF cells, respectively) were calculated for three randomly subsampled trials of two exemplars. Two of the three trials from each exemplar are used as training data, and the third as a test probe (this is done because the least-repeated exemplar has three trials). The pseudo-population response vectors are transformed into two-dimensional space using the classical multidimensional scaling algorithm in MATLAB (version 2018a, `cmdscale ()`), which linearly transforms the data into a two-dimensional space while optimally preserving Euclidean distance relationships between trials.

A Fisher's linear discriminant is fit to the training data, and the held-out test trials (one from each exemplar) are classified, resulting in an accuracy per iteration of 0, 0.5, or 1.0. This process was repeated for 10,000 iterations for each pair of exemplars, each time sampling a different set of trials, and the mean accuracy for each pair of categories, and within each category, for each iteration are compared with a surrogate distribution of 10,000 mean accuracies obtained by performing the same classification procedure on data with labels shuffled among the trials within the category (for within-category exemplar decoding) or among the categories (cross-category exemplar decoding) being compared. If the median rank exceeds 95% of the surrogate distribution, Bonferroni-corrected at $N = 15$ (for 5 within-category and 10 cross-category comparisons), the exemplars within the category or between categories are decoded above chance.

Multidimensional scaling plots. Log-firing rates from pseudo-populations of all 8 subjects in response to the first three house and face exemplars, normalized (*z*-scored) within each unit among trials, were transformed using the classical multidimensional scaling algorithm in MATLAB (`cmdscale ()`) to two-dimensional representations, first exemplars of both categories together, and then of houses and faces, alone.

Free recall preprocessing. Spike trains were aligned to stimulus onset. Firing rates for each candidate unit during key intervals (−0.1–0.1 s, 0.1–0.3 s, 0.3–0.5 s, and 0.5–1.6 s relative to image presentation) for face and place images were passed as input to a Friedman's test with trial number as the repeated measures factor. If $p < 0.025$ (Bonferroni-corrected at $N = 2$ for face and place categories), the unit was classified as visually responsive by the standard criterion; and if $p < 0.001$, it was classified as visually responsive by the strict criterion. A Wilcoxon rank-sum test was used to identify differences in firing rate between image categories on visually responsive unit spike trains 0.1-0.5 s after stimulus onset at $\alpha = 0.025$, and units with significant differences are classified as category selective.

Category preferences (face sensitivity, d') of units were calculated as follows:

$$d' = \sqrt{2}Z^{-1}\left(\frac{U}{n_1 n_2}\right)$$

where Z^{-1} is the inverse normal cumulative distribution function, U is the Mann–Whitney U statistic, and $n1$ and $n2$ are the number of place and face presentation trials, respectively. Face-selective units are redefined as those with a d' greater than that of the greatest d' associated with a non–category-selective unit by the rank-sum test (as above).

Free recall calculation of units' category preference. Spike trains were aligned to free recall speech. Mean firing rates during the time period between −2 and 2 s relative to utterance onset were calculated for each utterance and each unit, and the face sensitivity of spike rates associated with face and place utterances were calculated as above.

Selectivity of visually responsive units during free recall. The recall category preferences of units classified as visually responsive, according to the standard criterion, were compared with their respective category selectivities during image presentation. If the Spearman correlation coefficient was different from 0 at $\alpha = 0.05$, then the results were confirmed using only units that exceed the strict criterion for visual responsiveness.

For additional confirmation, the peri-recall time histogram (with bins of 0.1 s width) of each face-selective unit was divided by the mean firing rate of that unit over all recall periods. The baseline-corrected peri-recall time histograms for all face-selective units recorded for each recall event from each subject-implant were averaged together and convolved by a 1 s boxcar. (All face-selective units from each subject-implant were merged into a single "average unit" for that subject-implant, and the responses of these "average units" to place and face stimuli were compared at the last stage of analysis, as opposed to comparing places to faces for each unit, and then combining the differences. As a result, the decision to split one unit into two would not erroneously inflate significance in this calculation by increasing the number of units.) A two-tailed unpaired t test was used to compare mean activity −2 to 2 s relative to the onset of the set of face and place recollection utterances, respectively.

Repeated t tests were used to compare adjusted firing rates, downsampled to 2 Hz, at $\alpha = 0.05$, at each time point, without correction.

Changes in firing rate during face recall (−2 to 2 s relative to utterance onset compared with mean firing rate over face recall blocks) in each visually responsive unit (standard criterion) were compared with the changes during face image presentation (firing rate 0.1–0.5 s minus −0.4 to 0 s relative to image onset). Spearman's ρ and the parameters of the best fit line are calculated, and the correlation is tested against zero at $\alpha = 0.05$.

To identify reactivation in individual units, firing rate in the 2 s before face and place recall events were compared with the baseline rates for those respective visually responsive units at $\alpha = 0.05$ using a two-sample z test, with Bonferroni–Holm correction for multiple comparisons.

Relationship between firing rates during presentation and recall. Mean firing rates for each visually responsive (standard criterion) unit from each subject were calculated for −2 to 0 s relative to onset of face recall utterances. Mean firing rates from the same exemplars during presentation were weighted by the number of utterances in which they were referenced. Each visually responsive unit is plotted on a log-log scale based on presentation and recall firing rates. A first-degree polynomial is fit to the log-transformed data (0.01 pseudo-count added to avoid zeros), and R^2 fit is calculated, with the associated p value compared with 0.05.

Exemplar decoding of stimuli from free recall. The procedure for exemplar decoding in the 1-back task was repeated for data from the presentation phase of the free recall experiment, using mean firing rates between 0.1 and 0.5 s after image onset, to calculate decoding accuracy on an individual-subject level. Each exemplar was shown 4 times, so each classifier was trained on three trials of each exemplar and tested on one. Additionally, as shown in Figure 10, mean firing rates may differ between image presentation and recall, so to normalize each trial, the mean of each log-firing-rate pseudo-population vector was subtracted from each component of that vector, and all elements then divided by one greater than the SD (to avoid divide-by-zero errors). Category and cross-category accuracies were averaged across exemplar pairs and blocks at each iteration, and compared with a surrogate distribution obtained by shuffling the labels, as described for the 1-back task. The two runs performed by Subject 3 were coupled together as if they had occurred consecutively. A total of 10,000 cross-validation runs were performed, each time randomizing the test trial for each exemplar. The median percentile rank of the classifier accuracy for both categories and the cross-category result for each subject were compared with a statistical threshold of $\alpha = 0.05$, corrected for multiple comparisons using the Bonferroni–Holm technique ($N = 4$ subjects × 3 category domains), respectively. Only subjects and categories with significant presentation exemplar decoding were tested for recall exemplar decoding.

A population vector was built from each recall event, taking the log-base-10 (pseudo-count 1) of the mean firing rates of the neural ensemble from each subject, 2–0 s before the onset of the utterance. The population vector was normalized (as above) and classified with each of the binary classifiers trained to discriminate between each stimulus pair by presentation response (as above, but using all four trials for training). To calculate the cross-category classification performance, the full classifier array was applied to each recall datum; to calculate within-category classification performance, only the classifiers trained on pairs of faces or pairs of places were used on their respective stimuli.

The number of times that the classifiers from the array classified each input stimulus as belonging to a given exemplar is rank ordered. If a recall event is classified as belonging to the two different exemplars the same number of times, the ranking is repeated recursively using only the classifiers trained on those exemplars until the tie is broken or the number of classifications per exemplar stabilizes (in which case, the exemplars are assigned the mean of the two ranks). The ranks associated with the correct exemplar in all recall events for each stimulus category (faces and places, respectively) and all recall events together (cross-category) are added across blocks for each subject and compared with a surrogate distribution obtained by randomly permuting the labels of the recall events within each stimulus category, and repeating the classifier analysis 10,000 times (place and face) or 2000 times (cross-category). Surrogate distribution ranks of the accuracies of all subjects with significant category decoding during presentation were corrected for multiple comparisons using the Bonferroni–Holm method (among subject-categories with significant presentation exemplar decoding on the same task), and tested for significance at $\alpha = 0.05$.

To confirm successful recall exemplar decoding, exemplars were split into presentation response terciles for each face-selective unit, with the top and bottom terciles defined as the sets of exemplars evoking mean presentation responses above the 66.6th and below the 33.3rd percentiles of presentation mean response magnitudes (firing rates 0.1-0.5 s after presentation) across all four presentation trials, respectively. Exemplars were then split into recall activity terciles, with the top and bottom terciles defined as the sets of exemplars preceded by mean recall activity above the 66.6th and below the 33.3rd percentiles of mean recall activity magnitudes (mean firing rates 2–0 s before recall utterance) across all recall events associated with that exemplar. The mean, across face-selective units, of the fraction of exemplars in the top and bottom presentation response terciles that were also in the top and bottom recall activity terciles for each unit, respectively, relative to the total number of exemplars in the top and bottom recall activity terciles for that unit, was compared with a surrogate distribution obtained by pseudorandomly permuting the exemplar labels 10,000 times. The faction of matching exemplars so obtained was compared with $\alpha = 0.05$, with the Bonferroni–Holm correction for multiple comparisons (for the number subjects with face-selective units 3). In this context, more matching exemplars is equivalent to more congruence between presentation and recall activity for any given exemplar. This process was then repeated with only face exemplars in those subjects in whom congruence was above chance. Subject 6 was excluded because of a lack of face-selective units.

These analyses rely on shuffling labels of recall trials, so the quality of the surrogate distribution will inherently be limited by the number of recall events and recalled exemplars.

Results

Face-selective units found in VTC

Electrode positions were localized postoperatively by superimposing MRI and CT scans (Groppe et al., 2017); an example is

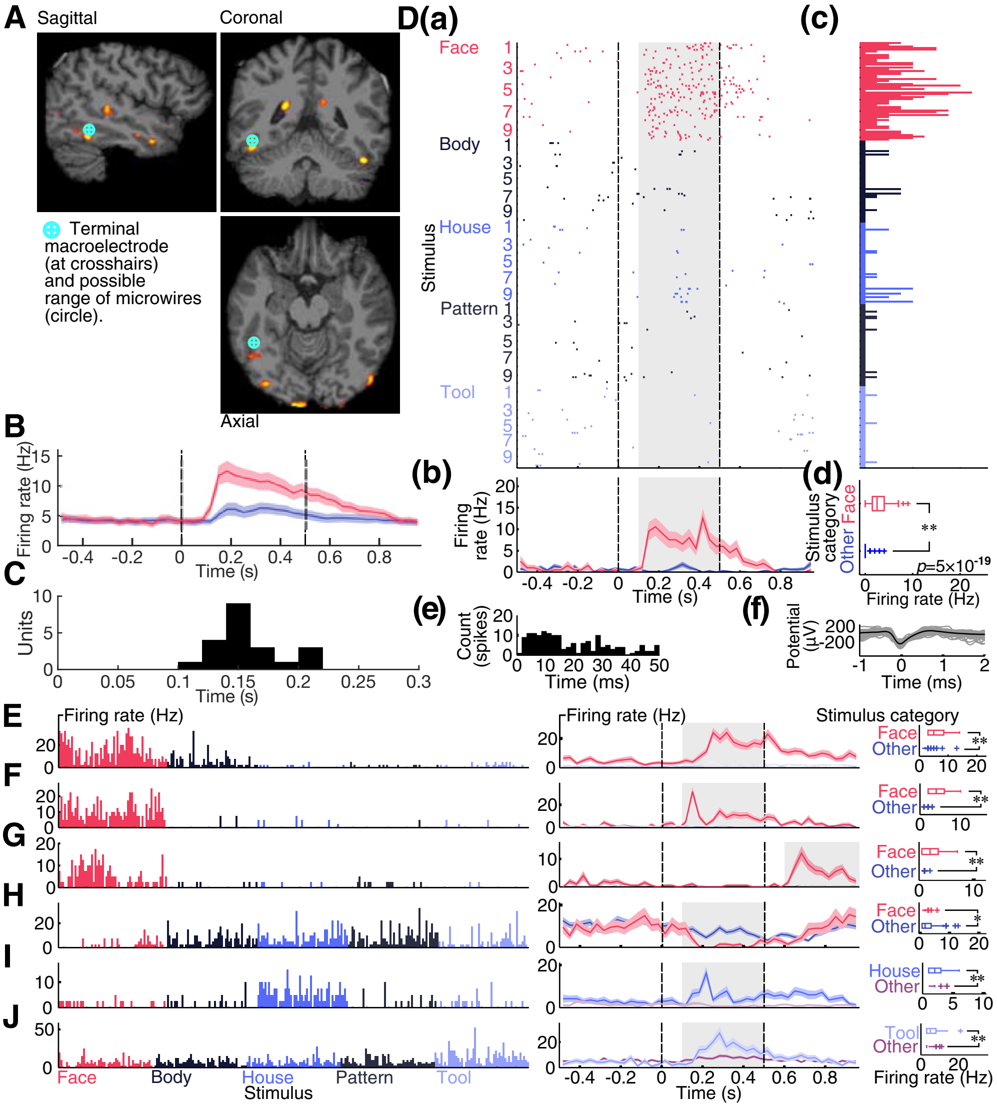

Figure 2. Diverse category-selective units found in the VTC. ***A***, Example electrode location (Subject 8) over fMRI face>house activation contrast map (in red, full data, Fig. 3, task structure: Fig. 1). ***B***, Grand mean peristimulus time histogram. Red represents faces. Blue represents nonface objects. Colored shaded areas represent mean ± SEM at each time point. Dashed lines indicate image onset at time 0 and offset, respectively. ***C***, Onset latency of face-selective units. ***D***, Representative face-selective unit from right FFA in Subject 8. ***Da***, Raster plot of responses with each row representing a single trial, and dashed lines indicating onset and offset, respectively. ***Db***, Mean peristimulus time histogram. Red represents faces. Blue represents nonface objects. Colored shaded areas represent mean ± SEM at each time point. ***Dc***, Mean firing rate per trial. Average over gray shaded area in raster plot, 0.1–0.5 s after presentation. ***Dd***, Distribution of responses to faces and nonface objects. Responses to faces are significantly stronger. ***De***, Interspike interval distribution. ***Df***, Spike waveforms. Black represents mean spike waveform. ***E–J***, Peristimulus time histograms and grand mean peristimulus time histograms for units with (***E***) longer response persistence, (***F***) transient peaked response, (***G***) offset response, (***H***) suppression to faces, (***I***) house selectivity, and (***J***) tool selectivity. Figure 4 shows distribution of unit selectivity in both tasks. Figure 5 shows positive correlations between responsiveness and face and category selectivity. $*p < 0.05$; $**p < 0.001$; rank-sum test, Bonferroni–Holm correction.

shown in Figure 2*A* (full data Fig. 3). None of the VTC areas sampled with microelectrodes were involved in seizure onset, as determined by epileptologist evaluation. Sixty-three visually responsive single units and multiunits were recorded collectively from 124 total units from all 8 subjects (for single-subject data, see Table 1). Of these, 33 were category-selective, with the vast majority (26, recorded from 5 subjects, 41% of visually responsive units) selective for faces (Fig. 4). All visually responsive units in Subject 1, the right electrode in Subject 3 and Subject 8 were face-selective. Among subjects with functional imaging, Subjects 3 (right electrode), 5, and 8 were the only ones to show electrode locations within reach of the FFA; Subject 1 has no functional imaging.

Thirty visually responsive units were not preferentially excited or inhibited by any specific category significantly more or less than all others, or were significantly excited and inhibited by two different categories, but neither to a significantly greater extent than the other. These units were prevalent in Subject 2, the left electrode in Subject 3, and in Subjects 4–7 (i.e., the Subject-electrodes that did not yield entirely face-selective units in this test). Thirty-two visually responsive units were single units and 31 were multiunits. The average time course of all visually responsive units across subjects showed a marked preference for faces (Fig. 2*B*), with onset latencies of face-selective units distributed between 100 and 220 ms after presentation of face stimuli (Fig. 2*C*). Visual responsiveness was strongly correlated with face selectivity (see Fig. 5; $p < 0.001$, Spearman's ρ, $N = 120$ units in subjects who performed the standard version of the task).

Figure 2*D* shows a typical face-selective single unit, with a vigorous response to face image presentation and return to baseline

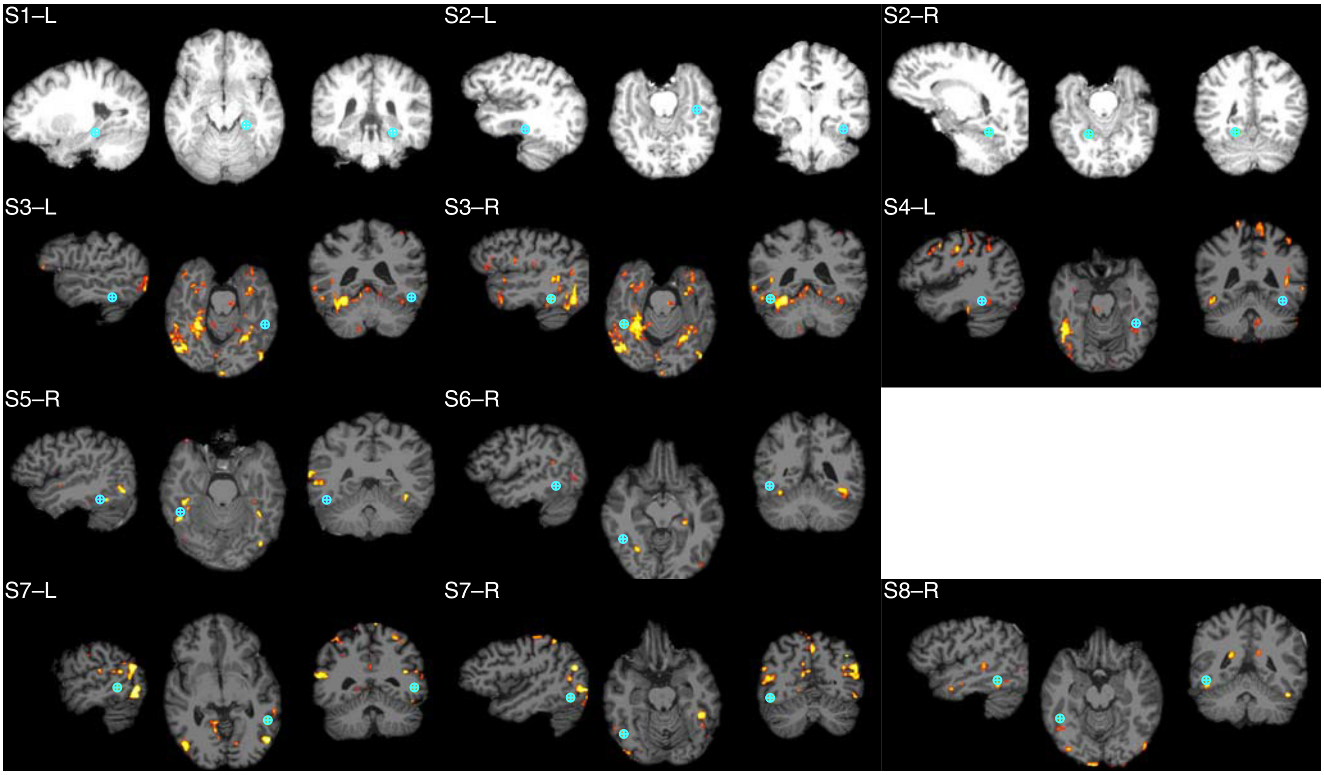

Figure 3. Electrode localizations from all subjects. S1-S8, Subjects 1-8; L, left electrodes; R, right electrodes. Crosshairs centered at terminal macro contact, with surrounding circles representing the 4 mm radius of the potential microelectrode range. Where functional imaging is available (Subjects 3-8), yellow/red areas represent regions with face>house responses, except in Subject 4, in which face>pattern responses were used.

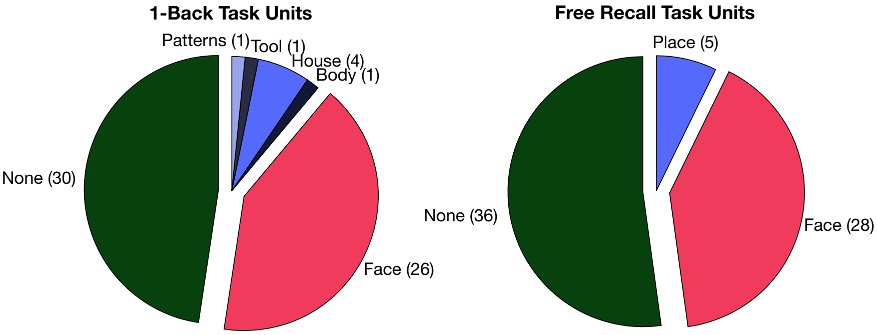

Figure 4. Pie chart showing the distributions of category selectivity among visually responsive units from all participating subjects in the 1-back and free recall/imagery tasks. The overwhelming majority of units that were selective for a single category were selective for faces.

firing rate after the image disappeared. However, we also observed a surprising diversity of face-selective response patterns, including units with response persistence (Fig. 2*E*) and units whose activity showed sharp transient peaks after face presentation (Fig. 2*F*). One single unit showed a strong face-selective offset response (Fig. 2*G*), but no significant onset response. Several units showed selective suppression to face presentation (Fig. 2*H*), a finding consistent with single-unit recordings from inferotemporal cortex in nonhuman primates (Gross et al., 1969; Freiwald and Tsao, 2010).

We also recorded units with selectivity for nonface categories: tools and houses. With the exception of two house-selective single units in 1 subject, from whom only weakly face-selective units were recorded, all house- and tool-selective units came from subjects whose recording sites yielded no face- or body-selective units. This is consistent with the reported segregation of domains dedicated to processing animate and inanimate objects (Chao et al., 1999; Weiner et al., 2014). After face-selective units, house-selective units were the most common, with four (6% of visually responsive units). Figure 2*I* shows one such unit. The sole tool-selective multiunit is shown in Figure 2*J*.

The one body-part-selective multiunit was inhibited by body part presentation. Our procedure also classified one unit as pattern-selective, but this was because patterns evoked a significantly lower-amplitude response in this unit relative to all other stimuli.

Using the alternative method for testing visual responsiveness (see Materials and Methods), at threshold $\alpha = 0.045$, 109 of 121 (90%) of units (excluding offset-responsive units) matched their responsiveness designation under the previously described regimen. Fifty-three of 62 visually responsive units identified by the rank-sum test were identified by the original method.

Exemplar decoding of faces in VTC ensembles

Next, we sought to corroborate previous reports suggesting that individual faces (face exemplars) had unique representations in the FFA (Nestor et al., 2011; Anzellotti et al., 2014; Axelrod and Yovel, 2015), in human VTC more broadly (Davidesco et al., 2014), and, at the single-unit level, in macaque face patches (Tsao et al., 2006, 2008;

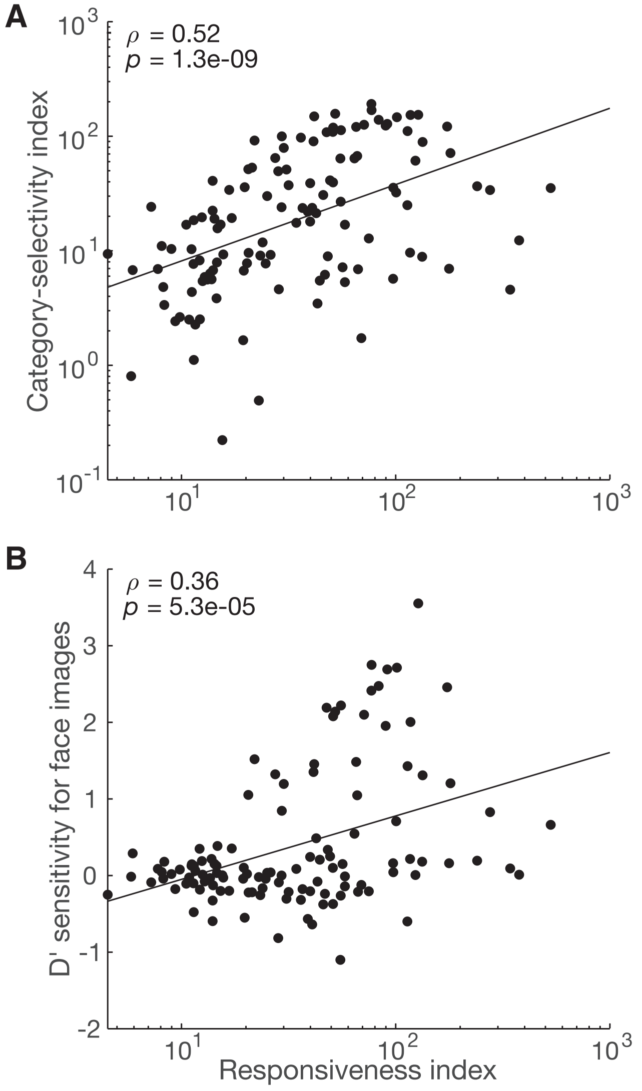

Figure 5. Correlation between unit visual responsiveness and (*A*) category selectivity, (*B*) d' face sensitivity over other categories. Visual responsiveness and category preference are strongly correlated ($p < 0.001$, Spearman's ρ, $N = 120$ units in subjects who performed the standard version of the task).

Chang and Tsao, 2017). Responses to single presentations (trials) of exemplars from all visually responsive units across all subjects were concatenated into a single log-transformed pseudo-population vector. Multidimensional scaling, a linear technique for dimensionality reduction, was then applied to visualize the relationships among trials of different exemplars in a common space. For example, responses from all trials of three selected face and house exemplars (Face 1–3 and House 1–3 in the stimulus set, chosen *ex ante*, for illustration, from the full set of 10 face and 10 house exemplars shown to each subject) are presented in Figure 6*B*. As expected, we observed a clear segregation of faces and houses in the representational space. We then applied multidimensional scaling to the three face exemplars (Fig. 6*C*) and the three house exemplars (Fig. 6*D*), alone, and found that trials of individual face exemplars appeared to be linearly separable, while trials of house exemplars were not.

To quantify this, we plotted transformed pseudo-population response vectors from each pair of exemplars presented (of 50, for a total of 1225 unique pairs) and used a simple linear discriminant to separate them. We then performed leave-one-out cross-validation as a measure of the discriminability of each of these exemplar pairs (Fig. 6*A*). Classifier accuracy was very high for exemplar pairs from different stimulus categories, especially faces versus nonface objects, showing that responses to these categories are distinct ($p < 0.0001$, random permutation test, Bonferroni correction at $N = 15$ category pairs). While it is unsurprising that faces can be distinguished from nonfaces, given the vastly different magnitudes of observed unit responses, the classifier was also able to discriminate between nonface categories, for example, tools and houses. Confirming the previous

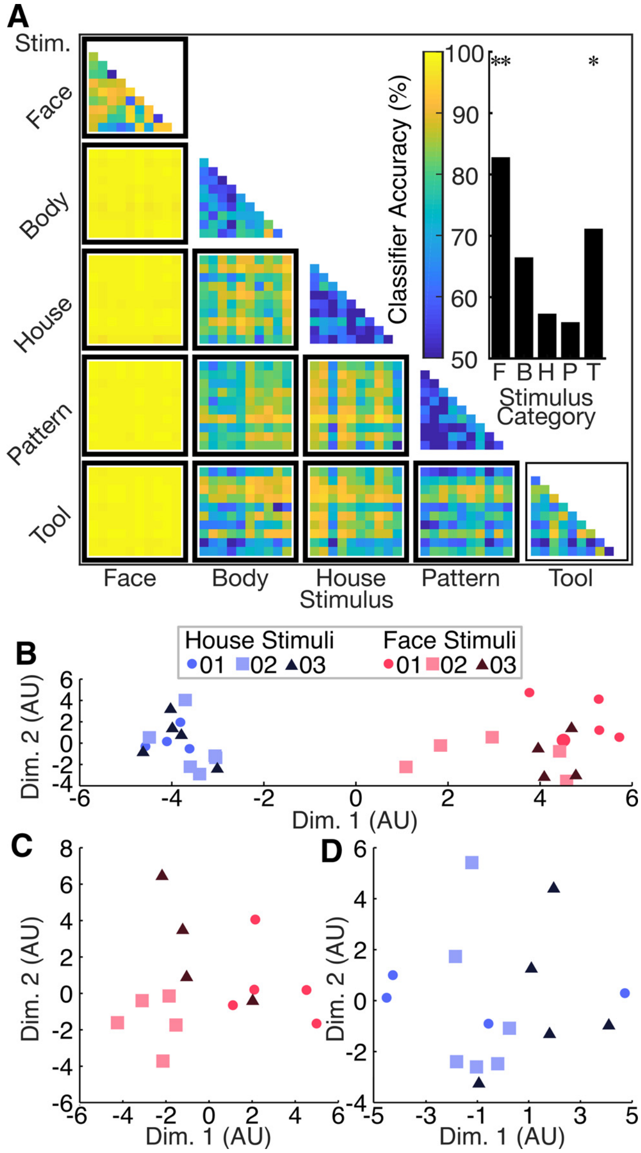

Figure 6. Exemplar decoding of face stimuli. ***A***, Binary classifier performance. Cross-validation accuracy of a set of classifiers tested and trained on each pair of exemplars. Discriminability and classifier accuracy of any exemplar pair represented by the color at that location in the matrix, with greater accuracy corresponding to more dissimilar population responses. Inset, Mean decoding accuracies for each within-category block (classifier accuracy discriminating exemplars of the same category). */thin box outline: $p < 0.05$; **/thick box outline: $p < 0.01$, random permutation test, Bonferroni correction. ***B***, Multidimensional scaling: population responses to trials of first three face and house exemplars. ***C***, Multidimensional scaling: population responses to trials of first three face exemplars. ***D***, Multidimensional scaling: population responses to trials of first three house exemplars.

studies, we show robust exemplar selectivity, evidenced by strong classifier performance in discriminating face exemplar pairs (83%, $p = 0.0001$, random permutation test, Bonferroni correction at $N = 5$ categories). We also found some weaker tool exemplar decoding (73%, $p = 0.008$, random permutation test, Bonferroni correction at $N = 5$ categories), but no within-category exemplar decoding for other categories (bodies: 66%, $p = 0.03$; houses: 57%, $p = 0.2$; patterns: 56%, $p = 0.3$, random permutation test, Bonferroni correction at $N = 5$ categories).

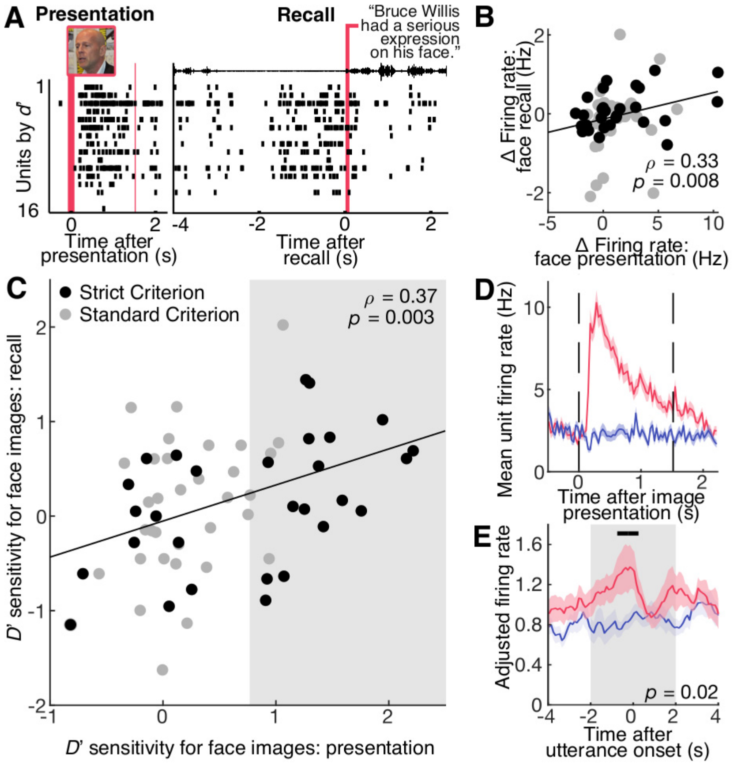

Figure 7. Face-selective units reactivate during recall. ***A***, Raster plot from presentation and recall example face stimulus from Subject 8. Image presentation of actor Bruce Willis† at time 0 on left, and vocalization onset (black envelope trace at top) on the right. Full run shown in Figure 8. ***B***, Responses to faces during presentation and before recall. Black represents units passing stricter visual responsiveness criterion. ***C***, Selectivity for faces during presentation and recall. Trend line fitted to all units. ***D***, Mean peristimulus time histogram for face trials. Face selectivity defined by gray area in ***C***. Red represents responses to faces. Blue represents responses to places. Colored areas represent mean ± SEM at each time point across subjects and trials ($N_1 = N_2 = 140$ place and face presentation trials). Dashed lines indicate image onset at time 0 and offset. ***E***, Mean peri-recall time histogram. Significant difference between face and place activity (gray box, −2 to 2 s, two-tailed unpaired t test, $p = 0.02$, $t_{(48)} = 2.4$, $N_1 = 24$ face recall events, $N_2 = 26$ place recall events). Figure 10 shows the relationships between mean presentation and recall firing rates among visually selective units. Black bar represents two-tailed unpaired t tests ($p < 0.05$, $t_{(48)} > 2.01$, uncorrected). †Image shown to subject is protected by copyright. Image in figure is a substitute. Image cropped from "Actors Helen Mirren and Bruce Willis," by Cameron Yee. © Cameron Yee. Licensed under Creative Commons Attribution 2.0 Generic (CC BY 2.0). The licensor does not endorse this work.

Free recall- and imagery-evoked reactivation of face representations

Next, we tested whether face representations in VTC could be activated endogenously in the absence of external visual input. Specifically, we wanted to learn whether VTC units that responded selectively to viewing of face images would respond in a similar way when subjects recalled or imagined those same face images. To that end, 4 subjects also performed an episodic free recall task (Norman et al., 2017), in which they were shown and asked to remember full-color photographs of famous faces and scenes. After performing a short interference task and putting a blindfold on, the subjects were asked to freely recall as many pictures as possible, focusing on one category (faces/places) at a time. The subjects were instructed to visualize and describe each picture they recalled in as much detail as possible, emphasizing unique colors, facial expressions, lighting, perspective, etc. Subjects 3, 6, and 8 recalled five, eight, and eight face exemplars and seven, ten, and five place exemplars, respectively. Subject 2 recalled three face and two place exemplars, and seven face and two place exemplars on the first and second runs, respectively. Face-selective units were recoded from 3 of 4 of this subset of subjects, and small numbers of place-selective units were recorded from 2 of 4 (also see Fig. 9A).

Guided by prior fMRI results (O'Craven and Kanwisher, 2000; Ishai et al., 2002) and bolstered by single-unit findings in the medial temporal cortex, (Kreiman et al., 2000) we first sought to identify whether face-selective units were reactivated during visual imagery of faces. We examined mean firing rates of VTC units within 2 s of the verbal recall event, reasoning that activity associated with visualization most likely occurred in this window. Units' firing rates during face presentation (over prestimulus baseline, see Methods and Materials), and in the 4 s interval centered at onset of the face recall utterance (over whole-experiment baseline), were well correlated ($p = 0.008$, Spearman's $\rho = 0.33$, $N = 62$ units; Fig. 7B), as were their preferences for face stimuli during presentation and recall ($p = 0.003$, Spearman's $\rho = 0.37$, $N = 62$ units; Fig. 7C). This correlation persisted when only strongly visually responsive units (see Methods and Materials) were included ($p = 0.002$, Spearman's $\rho = 0.55$, $N = 29$ units). To further examine this content-selective relationship, and to investigate the precise temporal dynamics of this recall-triggered activity, we computed the average baseline-corrected activity for each presentation trial (Fig. 7D) and each recall event (Fig. 7E) for all face-selective units in each implant, and compared the face to place stimuli. Activity in face-selective units was significantly greater around face recall events than around place recall events (unpaired t test, $p = 0.02$). Mean activity in face-selective units began increasing at ~2 s before onset of a face recall utterance, peaked, and returned to near baseline as the subject began to speak. This temporal relationship is consistent with fMRI (Polyn et al., 2005) results that demonstrate activation of face-selective brain areas, including the FFA, during recall, and iEEG results (Norman et al., 2017) showing an increase in the high-frequency broadband signal in category-selective VTC in the seconds leading up to recall of an item in that category. Subjects 2, 3, 6, and 8 had one, two, zero, and six units with significant activations in the 2 s before face recall (two-sample z

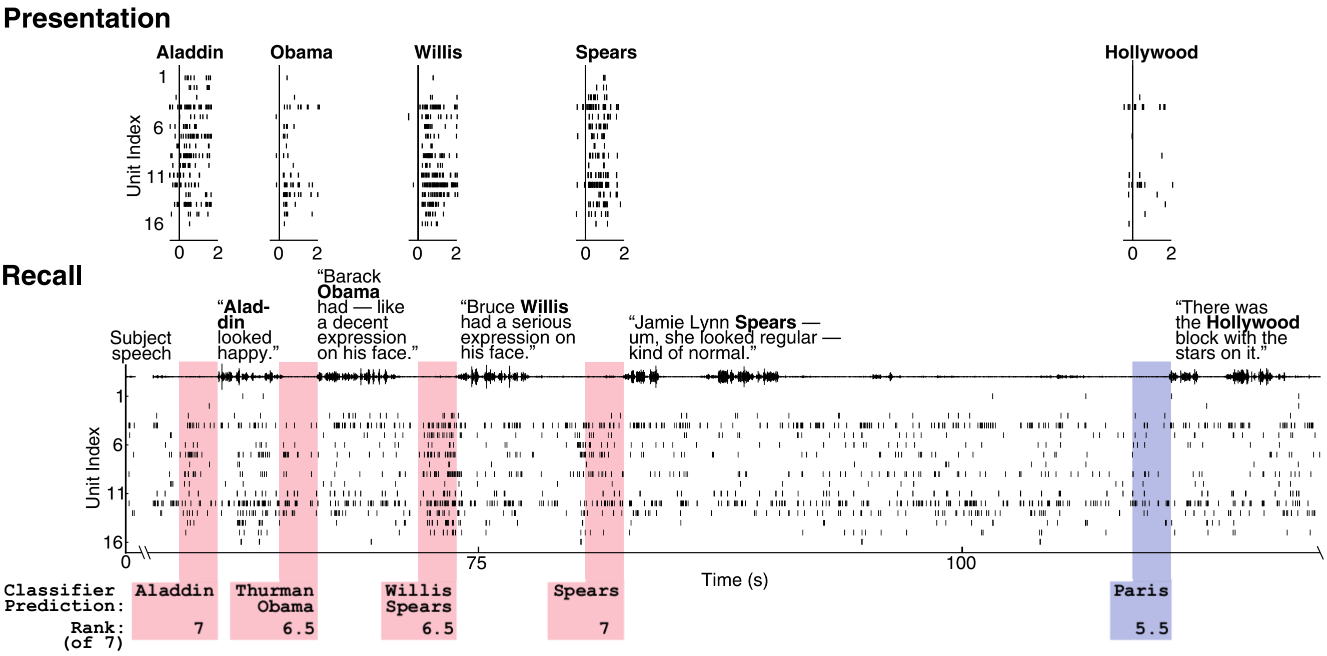

Figure 8. Audio and neural spike raster plot from an example face recall run from Subject 8, with associated presentation raster plots. Excerpt shows four face recall events (red) and one place recall event (blue), with speech audio envelope waveform and raster plot of 16 visually responsive and face-selective units. Raster plot from an example presentation of each exemplar, shown earlier in the experiment, above each respective recall event. Colored boxes highlight the 2 s before onset of each recall utterance, unit firing rates, which serve as the input for the classifier, and show the (top) classifier prediction (or top two predications if tied for first), and (bottom) the rank of the correct exemplar when voted on by the full set of classifiers, out of 7, where 7 is the best performance. Excerpts of subject speech with face or place identity in bold. The four face exemplars recalled by the subject rank either as the top classifier-set prediction, or are tied for first. The place exemplar is not predicted.

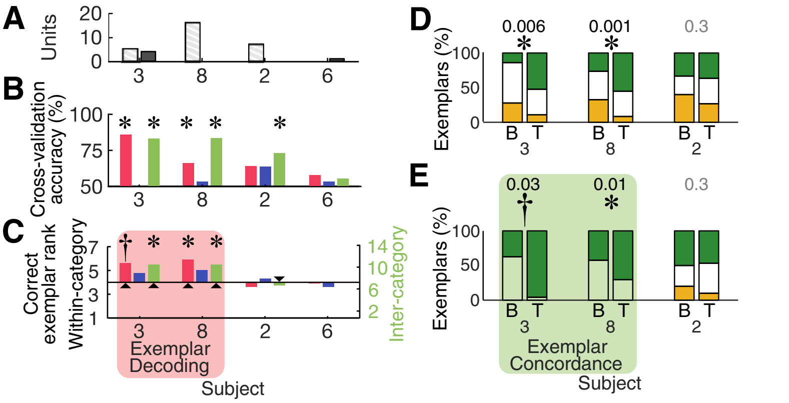

Figure 9. Exemplar decoding of recalled stimuli. ***A***, Number of face- (light, hashed bars) and place- (dark bars) selective units recorded in each subject in whom the experiment was performed. ***B***, Presentation exemplar decoding in recall task data. Red bars represent faces. Blue bars represent places. Green bars represent cross-category decoding. $*p < 0.05$ (random permutation test, Bonferroni–Holm correction). Exact *p* values in Table 2. ***C***, Recall exemplar decoding. Left axis: Red bars represent faces. Blue bars represent places. Green bars represent cross-category decoding. Right axis: Arrowheads indicate subject categories with presentation exemplar decoding. Red shaded subjects represent subjects with face presentation exemplar decoding (in whom face recall exemplar decoding is expected). $*p < 0.05$ (random permutation test, Bonferroni–Holm correction in subject categories with significant presentation exemplar decoding). †Trend toward significance, as preceding, but insufficient sample size on surrogate distribution. Exact *p* values in Table 3. ***D***, Mean concordance between presentation response and recall activity terciles among all exemplars, across face-selective units. B, Exemplars in bottom tercile of recall activity; T, top tercile. Green fraction represents exemplars in top tercile of presentation response. Yellow fraction represents bottom tercile. $*p < 0.05$ (random permutation test, Bonferroni–Holm correction). †Trend toward significance, as preceding, but limited sample size on surrogate distribution. Numbers above bars indicate uncorrected *p* value. ***E***, Mean concordance between presentation response and recall activity terciles among face exemplars, across face-selective units. Green shaded subjects represent subjects showing concordance among all exemplars (in whom face concordance is expected); otherwise, labeling as above.

test, Bonferroni–Holm correction, $N = 62$ units); all were face-selective during presentation, and none showed significant activation during place recall. Subject 6 had no face-selective units during presentation and, likewise, had no units with significant activations during face recall; Subject 6 also had the only unit with significant activation during place recall.

Next, we sought to determine whether neural ensembles from subjects with selectivity for individual face exemplars would demonstrate similar selectivity during recall. Figure 7*A* (and Fig. 8) shows single-trial raster plots for face-selective units during stimulation and recall. Patterns of activity before recall were strikingly similar to those observed during initial face presentation, suggesting that it might be possible to predict the identity of the face to be recalled by matching activity of these units before a recall event to their activity during presentation. To examine this possibility, we first identified all subjects whose recorded units showed above-chance exemplar decoding during presentation for each category. This reproduced the findings from the 1-back task (Fig. 6*A*) for each subject (Fig. 9*B*), with the exception that each pseudo-population vector was first normalized to account for the anticipated potential differences in magnitude of activation between presentation and recall periods (see Methods and Materials; Fig. 10). Mean leave-

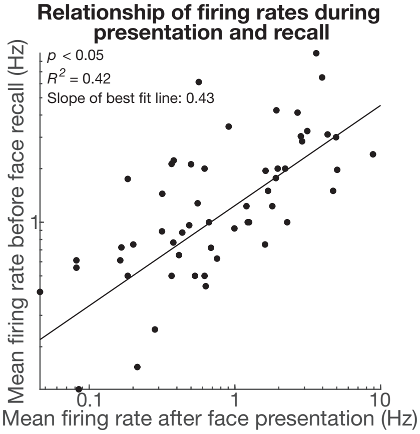

Figure 10. Mean firing rates during recall do not necessarily match those during presentation. Weighted mean of firing rates during presentation (0.1–0.5 s after onset) of remembered faces is plotted against mean firing rate before recall (−2 to 0 s relative to start of utterance) for visually responsive units ($N = 62$). For units that show strong activity during face presentation, average activity during face recall appears to be modestly weaker. This justifies using normalized firing rates as input to the classifiers trained on free recall and imagery task data. It does not imply that peak instantaneous firing rates are greater during presentation than recall. The slope is significantly different from zero ($p < 0.05$), as expected.

Table 2. *p* values associated with Figure 9*B*

	Subject 8	Subject 2	Subject 3	Subject 6
Faces	0.006	0.019	<0.001	0.12
Places	0.3	0.019	0.5	0.3
Intercategory	<0.001	<0.001	<0.001	0.09

Table 3. *p* values associated with Figure 9*C*

Subject	Subject 8	Subject 2	Subject 3	Subject 6
Faces	0.018	0.6	0.04	0.9
Places	0.07	0.17	0.18	0.4
Intercategory	<0.005	0.5	0.009	0.6

one-out cross-validation accuracy was calculated for classification among exemplars of the same category ("within-category") and all exemplars ("cross-category"). All subjects exhibiting face selectivity (Subjects 2, 3, and 8) showed above-chance cross-category classification accuracy (Fig. 9*B*, random permutation test, $p < 0.05$, Bonferroni–Holm correction at $N = 4$ subjects; for exact *p* values, see Table 2). Subjects 3 and 8 also showed above-chance within-category face classification accuracy. Subject 6 had no face-selective units (Fig. 9*A*) and, predictably, no exemplar decoding.

For categories that showed above chance decoding accuracy (Fig. 9*C*, arrowheads), we then tested whether the same classifiers, trained on the presentation data, would be able to decode exemplar identity based on activation before a recall event (i.e., cross-classification analysis). Subjects 3 and 8 (the 2 who showed exemplar decoding during presentation) showed significant exemplar decoding during recall for faces and across categories in combination ($p = 0.03$, cross-category and $p = 0.0017$ for faces, only, random permutation test, Fisher's method of combined probabilities); Subject 8 showed above-chance exemplar decoding individually; Subject 3 only recalled five face exemplars so the permutation test is not sufficiently powered to make strong claims of significance, but this subject nevertheless showed an accuracy value that would have satisfied the nominal significance criterion on this limited distribution ($p < 0.05$, random permutation test, Bonferroni–Holm correction at $N = 2$ and 3 tested subjects for faces and cross-category, respectively; for exact *p* values, see Table 3). Predictably, the excluded subject that did not demonstrate exemplar decoding during stimulation (Subject 2) also did not demonstrate it during recall.

To further confirm that face-selective units showed a consistent exemplar tuning between presentation and recall, we examined the proportion of exemplars in the top and bottom terciles of recall activity amplitude (measured using mean firing rates 2–0 s before recall utterance) for each face-selective unit that also evoked presentation responses in the top and bottom terciles, respectively, for that unit. Units with a large proportion of exemplars in the top and bottom terciles during recall that were also ranked in those same terciles during presentation have a strong concordance between presentation and recall tuning. As with the cross-classification analysis, both Subjects 8 and 3 (the 2 who showed exemplar decoding during presentation) showed above-chance concordance among face-selective units, among exemplars of both categories ("cross-category concordance"; Fig. 9*D*, $p < 0.05$, random permutation test, Bonferroni–Holm correction at $N = 3$ subjects). We then tested for concordance among only face exemplars. Again, the same subjects (8 and 3) showed significant concordance and a trend toward concordance (showed a concordance value that would have satisfied the nominal significance criterion had there been enough recall event data to fully define the surrogate distribution), respectively, between presentation and recall among face exemplars (Fig. 9*E*, $p < 0.05$, random permutation test, Bonferroni–Holm correction at $N = 2$ subjects with significant cross-category concordance).

As noted above, resolution of surrogate distributions for random permutation tests assessing recall exemplar decoding is limited by the number of exemplars recalled by each subject. For example, in the tercile split analysis, this translates to 30,240, 120, and 40,320 unique samples for Subjects 2, 3, and 8, respectively, for face exemplars (at least 5×10^8 unique permutations in each subject for all exemplars).

Discussion

We present the first comprehensive study of single-unit firing properties of the human VTC. Units in the vicinity of the FFA show a diverse range of responses to visual stimulation, including transient, sustained, suppressed, and offset responses. While suppression from baseline firing has been documented in area IT in nonhuman primates (Gross et al., 1972; Freiwald and Tsao, 2010), we are not aware of any reports of offset responses in higher visual areas of any animal model. In addition to face-selective units, we also recorded several units responsive to houses and tools in 3 subjects. In all but one of these (Subject 6), no face- or body-selective units were observed. In Subject 6, only weakly face-selective units were recorded. This is consistent with the previously reported segregation of animate and inanimate object representations in VTC (Chao et al., 1999; Weiner et al., 2014). All visually responsive units in

Subjects 1, 3 (right electrode), and 8 were face-selective, corresponding roughly to the electrodes whose localizations placed them within reach of the FFA: Subjects 3 (right electrode), 5, and 8; Subject 1 did not have functional imaging.

Next, we demonstrated that representations of individual face exemplars were consistent across trials. Individual face exemplars were strongly discriminable, an observation that corroborates previous reports of exemplar specificity in VTC that used high-frequency broadband activity as an approximate index of local population firing rate and that also showed that perceptual similarity among faces correlated to similarity in neuronal representation (Davidesco et al., 2014; Grossman et al., 2019). This is consistent with evidence from single-unit studies in nonhuman primates (Leopold et al., 2006; Freiwald et al., 2009; Chang and Tsao, 2017), a mechanistic understanding of these representations at the single-unit level is left to future investigators, who may choose to include perceptual similarity analysis of presented faces or a systematic survey of the face parameter space in their procedures.

At least one category-selective unit was recorded for tool, body part, and pattern categories each, but only tools showed evidence of exemplar decoding. Since we did not systematically record from areas selective for nonface categories and because so few of these units were recorded, we refrain from making claims about nonface exemplar coding or about whether differences in exemplar decoding between categories are significant based on the results of the present study.

Is the single-neuronal activity driven solely by the external visual stimulus, or is it also linked to the subjective, perceptual aspects of faces? To examine this, subjects were asked to report on, and visualize, freely recalled images, and we compared the associated recall-triggered responses in the purported face area to those evoked when subjects initially viewed those same images. Such experimental manipulation enabled us to assess the functional specificity of single neurons from the perceptual perspective, independent of the confounding influence of the physical, optical input, and to provide further empirical support for the claim that face representations in the VTC were related to perception rather than merely visually responsive. In general, units' face selectivity around recall events was correlated to their selectivity around presentation, and a mean of the activity of face-selective units across face and place recall events (Fig. 7*E*) showed that this was driven by a face-specific transient increase in face unit activity in the 2 s preceding recall (this analysis minimizes the influence of clustering decisions by first collapsing across all face-selective units in each subject). It was these 2 s of data that were subjected to the recall exemplar decoding analysis.

Subject 8 and, to some extent Subject 3, who showed exemplar decoding during presentation (i.e., a classifier was able to discriminate population codes for specific face exemplars when these subjects viewed those exemplars), also showed evidence of exemplar decoding during recall. In addition to yielding solely face-selective visually responsive units in the 1-back task, these subjects had electrodes in positions consistent with sampling the FFA. Crucially, these subjects showed exemplar decoding at the scale of millimeters of cortex (radius ~4 mm), whereas previous human studies (Davidesco et al., 2014) examined exemplar decoding over only broadly distributed areas (iEEG with inter-contact spacing of 0.5-1 cm). This means that exemplar-specific information is encoded in local areas within VTC, at the level of small neuronal ensembles, and is not just a feature of large neuronal populations. Face exemplars could feasibly be encoded by relatively limited, local integration of information broadly distributed across early visual areas.

There are several important caveats to this finding: Subject 3 recalled relatively few face exemplars (five), and this limits the number of unique permutations in the surrogate distribution used to test significance to only 120, potentially too few to claim $p < 0.05$ with adequate precision. The finding of face exemplar decoding in Subject 8 stands alone, but even 2 subjects would be too few to draw broad conclusions about the face coding in the human population. We present these limited results as they may still be of interest; more data will be required to elucidate the mechanisms of face coding in the human VTC. We also cannot make strong claims of causality at this time. Finally, our findings should not be interpreted as proof of face identity coding in the FFA, as we did not test multiple views of the same face identities; there is some evidence that more anterior VTC areas code for face identity in humans (Rajimehr et al., 2009; Anzellotti et al., 2014), but we do not test this here.

Subjects 2 and 6 also performed the free recall experiment, but recordings from Subject 6 yielded no face-selective units at that time, suggesting that the electrode was outside of a core face processing region. Subject 2 had many units that were not selective for faces in the previous 1-back task, and even one unit selective for body parts, suggesting that it, too, was outside of the FFA. In contrast, electrodes in Subjects 3 and 8 recorded only face-selective units in the 1-back task, analogous to recordings in monkey middle face patch, which also yield only face-selective cells (Tsao et al., 2006).

Regardless of the reason, Subjects 2 and 6 act as negative controls; if no exemplar decoding was found during presentation, none would be expected during recall. In contrast, subjects with exemplar decoding during presentation should exhibit such decoding during recall. This is the pattern we observed.

In a single-unit study of the medial temporal lobe in human subjects, Gelbard-Sagiv et al. (2008) were able to demonstrate that free recall of previously viewed video clips was preceded by content-specific reactivation of hippocampal and entorhinal neurons. Interestingly, such memory reactivation anticipated verbal responses by 1–2 s. More recently, we showed that population-level reactivations of higher-order visual areas during a free recall of images (as measured by high-frequency broadband iEEG activity) occurred contemporaneously with sharp-wave ripples in the hippocampi of human subjects in a manner that matched local category preferences and exemplar-specific activation patterns from the presentation phase of the experiment in the same iEEG electrodes. The frequency of these sharp-wave ripple events increased significantly in the 1–2 s before a free recall event, matching the chronology seen in the current data for unit-level reactivation (Norman et al., 2019).

Despite limitations, our findings illustrate convincingly that face-selective units are recruited in category-specific recall and/or imagery. This builds on a long line of literature in nonhuman primates showing that neurons in higher visual areas respond to top-down influences, including working memory for colors (Fuster and Jervey, 1982) and fractal shapes (Miyashita and Chang, 1988). In humans, category-specific visual areas have been shown to reactivate to imagery, including reactivation of face areas to face recall and imagery (O'Craven and Kanwisher, 2000; Ishai et al., 2002; Norman et al., 2017). We have previously shown that reinstatements during free recall are associated with sharp wave ripples in hippocampus (Norman et al., 2019); however, further research is needed to characterize the representational content that is being reinstated (i.e., low-level pictorial representations or high-level semantic features).

In conclusion, we present a detailed study of single units of the human VTC, demonstrating properties of responsiveness to face and nonface stimuli during viewing and during an imagery task. Extending prior observations from nonhuman primates, we report a diverse range of highly face-selective units within human VTC; that those units form a population code by which individual face exemplars can be discriminated; and that reactivation of the patterns forming that code occurs not only during face perception, but also perhaps during face imagination and recall. In line with prior neuroimaging work supporting a role of the VTC in conscious perception, we demonstrate selective activation during recall at the single neuron level. These findings support the role of the VTC, and FFA specifically, as a critical substrate of conscious face representation, one used not only to identify and discriminate faces in the environment, but also to support face representations generated internally. Our research adds to a large and growing body of literature supporting a role for higher-order sensory areas in subserving working memory, imagery, and other cognitive processes that engage the same neuronal substrates as bottom-up sensory processes.

References

Afraz A, Boyden ES, DiCarlo JJ (2015) Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. Proc Natl Acad Sci USA 112:6730–6735.

Anzellotti S, Fairhall SL, Caramazza A (2014) Decoding representations of face identity that are tolerant to rotation. Cereb Cortex 24:1988–1995.

Axelrod V, Rozier C, Malkinson TS, Lehongre K, Adam C, Lambrecq V, Navarro V, Naccache L (2019) Face-selective neurons in the vicinity of the human fusiform face area. Neurology 92:197–198.

Axelrod V, Yovel G (2015) Successful decoding of famous faces in the fusiform face area. PLoS One 10:e0117126.

Bollinger J, Rubens MT, Zanto TP, Gazzaley A (2010) Expectation-driven changes in cortical functional connectivity influence working memory and long-term memory performance. J Neurosci 30:14399–14410.

Chang L, Tsao DY (2017) The code for facial identity in the primate brain. Cell 169:1013–1028.e14.

Chao LL, Haxby JV, Martin A (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat Neurosci 2:913–919.

Cowan N (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychol Bull 104:163–191.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Davidesco I, Zion-Golumbic E, Bickel S, Harel M, Groppe DM, Keller CJ, Schevon CA, McKhann GM, Goodman RR, Goelman G, Schroeder CE, Mehta AD, Malach R (2014) Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. Cereb Cortex 24:1879–1893.

Desimone R, Albright TD, Gross CG, Bruce C (1997) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:1–12.

DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? Neuron 73:415–434.

Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330:845–851.

Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. Nat Neurosci 12:1187–1196.

Fuster JM, Jervey JP (1982) Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. J Neurosci 2:361–375.

Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, Fried I (2008) Internally generated reactivation of single neurons in human hippocampus during free recall. Science 322:96–101.

Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48:63–72.

Groppe DM, Bickel S, Dykstra AR, Wang X, Mégevand P, Mercier MR, Lado FA, Mehta AD, Honey CJ (2017) iELVis: an open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. J Neurosci Methods 281:40–48.

Gross CG, Bender DB, Rocha-Miranda CE (1969) Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science 166:1303–1306.

Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophysiol 35:96–111.

Grossman S, Gaziv G, Yeagle EM, Harel M, Mégevand P, Groppe DM, Khuvis S, Herrero JL, Irani M, Mehta AD, Malach R (2019) Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. Nat Commun 10:13.

Ishai A, Haxby JV, Ungerleider LG (2002) Visual imagery of famous faces: effects of memory and attention revealed by fMRI. Neuroimage 17:1729–1741.

Jacques C, Witthoft N, Weiner KS, Foster BL, Rangarajan V, Hermes D, Miller KJ, Parvizi J, Grill-Spector K (2016) Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. Neuropsychologia 83:14–28.

Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5:143–156.

Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825–841.

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782–790.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kreiman G, Koch C, Fried I (2000) Category-specific visual responses of single neurons in the human medial temporal lobe. Nat Neurosci 3:946–953.

Leopold DA, Bondar IV, Giese MA (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442:572–575.

Liu J, Harris A, Kanwisher N (2002) Stages of processing in face perception: an MEG study. Nat Neurosci 5:910–916.

Minear M, Park DC (2004) A lifespan database of adult facial stimuli. Behav Res Methods Instrum Comput 36:630–633.

Misra A, Burke JF, Ramayya AG, Jacobs J, Sperling MR, Moxon KA, Kahana MJ, Evans JJ, Sharan AD (2014) Methods for implantation of micro-wire bundles and optimization of single/multi-unit recordings from human mesial temporal lobe. J Neural Eng 11:026013.

Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331:68–70.

Mossbridge JA, Grabowecky M, Paller KA, Suzuki S (2013) Neural activity tied to reading predicts individual differences in extended-text comprehension. Front Hum Neurosci 7:655.

Nestor A, Plaut DC, Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. Proc Natl Acad Sci USA 108:9998–10003.

Niediek J, Boström J, Elger CE, Mormann F (2016) Reliable analysis of single-unit recordings from the human brain under noisy conditions: tracking neurons over hours. PLoS One 11:e0166598.

Norman Y, Yeagle EM, Harel M, Mehta AD, Malach R (2017) Neuronal baseline shifts underlying boundary setting during free recall. Nat Commun 8:1301.

Norman Y, Yeagle EM, Khuvis S, Harel M, Mehta AD, Malach R (2019) Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. Science 365:eaax1030.

O'Craven KM, Kanwisher N (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J Cogn Neurosci 12:1013–1023.

Papademetris X, Jackowski MP, Rajeevan N, DiStasio M, Okuda H, Constable RT, Staib LH (2006) BioImage Suite: an integrated medical image analysis suite. An update. Insight J 2006:209.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005) Category-specific cortical activity precedes retrieval during memory search. Science 310:1963–1966.

Puri AM, Wojciulik E, Ranganath C (2009) Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. Brain Res 1301:89–99.

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. Nature 435:1102–1107.

Rajimehr R, Young JC, Tootell RB (2009) An anterior temporal face patch in human cortex, predicted by macaque maps. Proc Natl Acad Sci USA 106:1995–2000.

Ranganath C, Cohen MX, Dam C, D'Esposito M (2004) Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. J Neurosci 24:3917–3925.

Singer JM, Madsen JR, Anderson WS, Kreiman G (2015) Sensitivity to timing and order in human visual cortex. J Neurophysiol 113:1656–1669.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670–674.

Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. Proc Natl Acad Sci USA 105:19514–19519.

Weiner KS, Golarai G, Caspers J, Chuapoco MR, Mohlberg H, Zilles K, Amunts K, Grill-Spector K (2014) The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. Neuroimage 84:453–465.

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci USA 111:8619–8624.