Title: Face-selective units in human ventral temporal cortex reactivate during free recall

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

Affiliations:

[1]Department of Neurosurgery, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, and Feinstein Institute for Medical Research, 350 Community Dr., Manhasset, NY 11030, USA.

[2]Department of Neurobiology, Weizmann Institute of Science, 76100 Rehovot, Israel.

*Correspondence to: A.D.M. at amehta@northwell.edu.

Abstract:

Properties of face-responsive individual neurons in the human ventral temporal cortex (VTC) have yet to be studied, and their role in conscious perception remains unknown. To explore this, we implanted microelectrodes into the VTCs of eight human subjects undergoing invasive epilepsy monitoring. Most (26 of 33) category-selective units showed specificity for face stimuli, with a range of response profiles. Different face exemplars evoked consistent and discriminable responses in the population of units sampled. During a free recall task, face-selective units selectively reactivated in the absence of visual stimulation during the 2-second window prior to face recall events. Furthermore, the identity of the recalled face could be predicted by comparing activity preceding recall events to activity evoked by visual stimulation.

One Sentence Summary: Single neurons in the human ventral temporal cortex code for individual face exemplars, both during sensory stimulation and during imagery in the absence of sensory stimulation.

Main Text:

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 therein must meet social demands to recognize, classify and uniquely identify a multitude of faces. First described in monkeys, studies of single neuron firing in response to visual features of faces have uncovered key "bottom-up" mechanisms of the feature space that drives neuronal firing (*1–3*). VTC neurons exhibit precise facial feature sensitivity, supporting their role in the discrimination of individual face exemplars. However, their role in non-sensory, extraretinal processing, as would occur during imagery and recall, is difficult to extrapolate from monkeys who cannot qualitatively express their experience. In humans, neuroimaging has defined a critical node in face processing within the VTC: the fusiform face area (FFA) (*4*). Functional MRI (fMRI) studies support both a sensory ("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 (*5*, *6*), perform visual imagery tasks involving faces (*7*, *8*), and hold face representations in working memory (*9*). Activation of category-selective regions of VTC can predict recall of items in that category (*10*, *11*). In addition to fMRI studies, magnetoencephalography (*12*) and direct electrocortical recordings (*13*) in humans show

fusiform responses emerging early (~100 ms) after face presentation, which have been interpreted as evidence for a sensory role for this region. While these studies have provided important insights, they lack the spatiotemporal resolution needed to uncover response properties of individual human neurons. Thus, the exquisite 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 modified to include microwires provide an opportunity to study neuronal spiking in patients undergoing chronic recordings, and such *in vivo* human data have provided insights into a number of physiological and pathological processes, most notably in the medial temporal lobe (*14*). However, single and population spiking activity in the VTC has yet to be explored. To bridge these knowledge gaps, 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 as well as higher-order cognitive processes at the single-unit level. Recorded units exhibited a wide 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 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 episodic memory in which the single-neuronal substrate of sensory processing is reactivated in a top-down fashion during recollection (*15*).

Eight subjects (four females, see Table S1) with partial epilepsy, undergoing diagnostic intracranial EEG (iEEG) monitoring of VTC were implanted with micro-macro depth electrodes (*16*) (Ad-Tech Medical, Oak Creek, WI), using either anatomical landmarks (*17*) or fMRI guidance. All subjects gave informed consent under IRB guidelines, and electrode targeting was guided primarily by clinical needs. Electrode positions were localized post-operatively by superimposing MRI and computed tomography scans (*18*), and an example is shown in Fig. 1A (full data, Fig. S1). None of these areas were involved in seizure onset, as determined by epileptologist evaluation. Subjects performed a 1-back task with faces, body parts, houses, tools and patterns (Fig. S2) while microelectrode data were recorded at 30 kHz. Sixty-three visually-responsive single and multi-units were recorded collectively from all eight subjects (see Table S1 for single-subject data). Of these, 33 were category-selective, with the vast majority (26, recorded from five subjects, 41% of visually-responsive units) selective for faces (see Fig. S3). The average time-course of all visually responsive units across subjects showed a marked preference for faces (Fig. 1B). Visual responsiveness was strongly correlated with face selectivity ($p < 0.001$, Spearman's ρ, Fig. S4).

Fig. 1C shows a typical face-selective unit, with a vigorous response to face image presentation and return to baseline firing rate after the image disappeared. However, we also observed a surprising diversity of face-selective response patterns, including units with response persistence (Fig. 1D, Fig. S5) and units whose activity showed sharp transient peaks after face presentation (Fig. 1E, Fig. S6). One unit showed a strong face-selective offset response (Fig. 1F, Fig. S7). Several units showed selective suppression to face presentation (Fig. 1G, Fig. S8), a finding consistent with single-unit recordings from inferotemporal cortex in non-human primates (*19*, *20*).

We also recorded units with selectivity for non-face categories – tools and houses. With the exception of two house-selective units recorded from one 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 (*17*, *21*). After face-selective units, house-selective units were the most common, with four (6% of visually-responsive units). Fig. 1H shows one such unit (also see Fig. S9). The sole tool-selective unit is shown in Fig. 1I (also see Fig. S10).

Next, we sought to corroborate previous reports suggesting that individual faces (face exemplars) had unique representations in the FFA (*22–24*), in human VTC more broadly (*25*), and, at the single-unit level, in macaque face patches (*1–3*). 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. Multi-dimensional scaling (MDS), 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 face and house exemplars (Face 1–3 and House 1–3 in the stimulus set, chosen *ex ante*, for illustration) are presented in Fig. 2B. As expected, we observed a clear segregation of faces and houses in the representational space. We then applied MDS to the three face exemplars (Fig. 2C) and the three house exemplars (Fig. 2D), 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 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. 2A). Classifier accuracy was very high for exemplar pairs from different stimulus categories, especially faces *vs*. non-face objects, showing that responses to these categories are distinct ($p < 0.01$, bootstrap test, Bonferroni correction). While it is unsurprising that faces can be distinguished from non-faces, given the vastly different magnitudes of the responses of many of the units, the classifier was also able to distinguish among non-face categories. Confirming the previous studies, we show robust exemplar selectivity, evidenced by strong classifier performance in discriminating face exemplar pairs (>80%, $p < 0.01$, bootstrap test, Bonferroni correction). We also found some weaker tool exemplar decoding ($p < 0.05$, bootstrap test, Bonferroni correction), but no within-category exemplar decoding for other categories.

Next, we tested whether face representations could be activated spontaneously by the brain, during imagery in the absence of external visual input. To that end, half of the subjects also performed an episodic free recall task, reported before (*11*), 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. Importantly, the subjects were instructed to visualize and describe each picture that they recall in as much visual detail as possible, emphasizing unique colors, face expression, lighting, perspective, *etc*. Face-selective units were recoded from 3/4 subjects, and small numbers of place-selective units were recorded from 2/4 (Fig. S11A).

We observed that units' firing rates during face presentation (over pre-stimulus baseline), and in the 4-second interval centered at onset of the face recall utterance (over whole-experiment baseline), were well correlated ($p < 0.05$, Spearman's $\rho = 0.33$, Fig. 3C), as was their preference for face stimuli during presentation and recall ($p < 0.05$, Spearman's $\rho = 0.37$, Fig. 3D). These correlations persisted when only strongly visually-responsive units (see *Methods*) were included. To further examine this content-selective relationship, and to investigate the precise temporal

dynamic of this recall-triggered activity, we computed the average baseline-corrected activity for each presentation trial (Fig. 3E) and each recall event (Fig. 3F) 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 (2-sample *t*-test, $p < 0.05$). Mean activity in face-selective units began increasing around 2 seconds before onset of a face recall utterance, peaked, and returned to near baseline as the subject began to speak. This result is consistent with previous fMRI (*10*) and iEEG (*11*) studies showing an increase in blood-oxygenation-level dependent (BOLD) activity and high-frequency broadband signal in category-selective VTC in the seconds leading up to recall of an item in that category.

Single-trial raster plots (Fig. 3A, Fig. S12) for face-selective units displayed remarkably similar patterns of activity before recall and during the initial face image presentation, suggesting it might be possible to predict recall of individual face exemplars using firing patterns of these units during presentation. To do this, we trained a classifier similar to the one used to test for exemplar decoding in the 1-back task (Fig. 2) on data from stimulus presentation and attempted to classify individual recall events based on the activity of the units in the 2 seconds prior to utterance onset. As mean firing rates can vary between during presentation and recall (Fig. S13), we normalized across the population before the classifier was applied.

Leave-one-out cross-validation was performed on a set of binary classifiers trained to discriminate each pair of exemplars. Classification accuracy was defined as correctly identifying the exemplar from which the input trial was drawn (for within-category classification), and correctly identifying its category, regardless of the specific exemplar (in cross-category classification). Three out of the four subjects showed above-chance cross-category classification accuracy (bootstrap test, $p < 0.05$, Bonferroni-Holm correction, Fig. S11B), and two also showed above-chance face classification accuracy. This corresponds to the set of subjects from whom face-selective units were recorded, reinforcing that exemplar-selective face information can be decoded at the level of few neurons spatially confined to the sampling volume of a single microwire bundle, no more than 8 mm across.

We proceeded to test the performance of our classifier sets in decoding exemplar or category based on significant activations prior to recall utterances (Fig. 3B, S11). Subject 3 showed above-chance cross-category decoding ($p < 0.05$, bootstrap test, Bonferroni-Holm correction), while Subject 8 showed a similar trend. Taken together, Subjects 3 and 8 clearly showed face exemplar decoding above chance ($p < 0.05$, bootstrap test, Bonferroni-Holm correction, complete data Fig. S11C–D).

In summary, we demonstrate the first single unit recordings from the human VTC. 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 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 the FFA specifically, as the substrate of conscious face representation in the human brain, and that this structure is used not only to identify and discriminate faces observed in the environment, but also to host internally-generated representations. Our research adds to a large and growing body of literature supporting a multi-faceted role for higher-order sensory areas in

subserving working memory, imagery and other processes whereby higher-order cognitive events engage the same neuronal substrates as bottom-up sensory processes.

References and Notes:

1. D. Y. Tsao, W. A. Freiwald, R. B. H. Tootell, M. S. Livingstone, A cortical region consisting entirely of face-selective cells. *Science*. **311**, 670–4 (2006).
2. D. Y. Tsao, S. Moeller, W. A. Freiwald, Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci.* **105**, 19514–19519 (2008).
3. L. Chang, D. Y. Tsao, The Code for Facial Identity in the Primate Brain. *Cell*. **169**, 1013–1028.e14 (2017).
4. N. Kanwisher, J. McDermott, M. M. Chun, The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–11 (1997).
5. A. M. Puri, E. Wojciulik, C. Ranganath, Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Res.* **1301**, 89–99 (2009).
6. J. Bollinger, M. T. Rubens, T. P. Zanto, A. Gazzaley, Expectation-driven changes in cortical functional connectivity influence working memory and long-term memory performance. *J. Neurosci.* **30**, 14399–410 (2010).
7. A. Ishai, J. V Haxby, L. G. Ungerleider, Visual imagery of famous faces: Effects of memory and attention revealed by fMRI. *Neuroimage*. **17**, 1729–1741 (2002).
8. G. Ganis, W. L. Thompson, S. M. Kosslyn, Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cogn. Brain Res.* **20**, 226–241 (2004).
9. C. Ranganath, M. X. Cohen, C. Dam, M. D'Esposito, Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *J. Neurosci.* **24**, 3917–25 (2004).
10. S. M. Polyn, V. S. Natu, J. D. Cohen, K. A. Norman, Category-specific cortical activity precedes retrieval during memory search. *Science*. **310**, 1963–6 (2005).
11. Y. Norman, E. M. Yeagle, M. Harel, A. D. Mehta, R. Malach, Neuronal baseline shifts underlying boundary setting during free recall. *Nat. Commun.* **8**, 1301 (2017).
12. J. Liu, A. Harris, N. Kanwisher, Stages of processing in face perception: An MEG study. *Nat. Neurosci.* **5**, 910–916 (2002).
13. C. Jacques *et al.*, Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuropsychologia*. **83**, 14–28 (2016).
14. N. Suthana, I. Fried, Percepts to recollections: insights from single neuron recordings in the human brain. *Trends Cogn. Sci.* **16**, 427–36 (2012).
15. N. Cowan, Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull.* **104**, 163–91 (1988).
16. A. Misra *et al.*, in *Journal of Neural Engineering* (NIH Public Access, 2014; http://www.ncbi.nlm.nih.gov/pubmed/24608589), vol. 11, p. 026013.
17. K. S. Weiner *et al.*, The mid-fusiform sulcus: A landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage*. **84**, 453–465 (2014).
18. D. M. Groppe *et al.*, iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. *J. Neurosci. Methods*. **281**, 40–48 (2017).
19. C. G. Gross, D. B. Bender, C. E. Rocha-Miranda, Visual receptive fields of neurons in

inferotemporal cortex of the monkey. *Science*. **166**, 1303–6 (1969).
20. W. A. Freiwald, D. Y. Tsao, Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*. **330**, 845–51 (2010).
21. L. L. Chao, J. V. Haxby, A. Martin, Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.* **2**, 913–919 (1999).
22. A. Nestor, D. C. Plaut, M. Behrmann, Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc. Natl. Acad. Sci.* **108**, 9998–10003 (2011).
23. S. Anzellotti, S. L. Fairhall, A. Caramazza, Decoding representations of face identity that are tolerant to rotation. *Cereb. Cortex*. **24**, 1988–1995 (2014).
24. V. Axelrod, G. Yovel, Successful decoding of famous faces in the fusiform face area. *PLoS One*. **10**, e0117126 (2015).
25. I. Davidesco *et al.*, Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. *Cereb. Cortex*. **24**, 1879–1893 (2014).
26. M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, S. M. Smith, FSL. *Neuroimage*. **62**, 782–790 (2012).
27. M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. **17**, 825–841 (2002).
28. D. N. Greve, B. Fischl, Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*. **48**, 63–72 (2009).
29. M. Jenkinson, S. Smith, A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**, 143–156 (2001).
30. X. Papademetris *et al.*, BioImage Suite: An integrated medical image analysis suite: An update. *Insight J.* **2006**, 209 (2006).
31. M. Minear, D. C. Park, A lifespan database of adult facial stimuli. *Behav. Res. Methods, Instruments, Comput.* **36**, 630–633 (2004).
32. J. Niediek, J. Boström, C. E. Elger, F. Mormann, Reliable analysis of single-unit recordings from the human brain under noisy conditions: Tracking neurons over hours. *PLoS One*. **11**, e0166598 (2016).
33. 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. Copyright Cameron Yee. Licensed under Creative Commons Attribution 2.0 Generic (CC BY 2.0). The licensor does not endorse this work.

Acknowledgments: We thank Salman Qasim and Drs. Joshua Jacobs, Elliot Smith, Ella Podvalny, Tal Golan, Pierre Megévand, José Herrero, Victor Du and Stephan Bickel for intellectual and technical support. We thank Dr. Xiangzhi Zhou for help with the fMRI experiments. We acknowledge the contributions of our colleagues at the Northwell Health Comprehensive Epilepsy Center, especially Willie Walker, R. EEG. T., Monika Lalik, R. EEG. T., and Drs. Sean Hwang, Fred Lado and Scott Stevens. We would like to thank our patient-subjects and their families, without whose patience and cooperation, this research would have been impossible. **Funding:** A.D.M. receives funding from the following grants: NIH/NINDS NS098976-01, NSF-BSF-2017015; NIMH MH114166-01; S.K. is supported by the Feinstein M.D. Scholarship; This work was partly funded by US-Israel BSF grant to A.D.M. and R.M.; **Author contributions:** All authors contributed to conceptualization of this research. S.K.

performed formal analysis on electrophysiological data, and E.M.Y. and S.G. on imaging data. S.K. and E.M.Y. collected the data. S.K. and Y.N. developed the methodology. A.D.M. performed all of the surgeries. A.D.M. and R.M. provided resources and supervision. S.K. and E.M.Y. created visualizations. S.K. and A.D.M wrote the original draft; all authors reviewed and edited the manuscript. **Competing interests:** A.D.M. has a consulting agreement with Medtronic. **Data and materials availability:** Data will be provided upon request. Code is available on: https://github.com/IEEG/SUFreeRecall.

Supplementary Materials:

Materials and Methods

Supplementary Text

Figures S1-S13

Table S1

References (*26-32*)

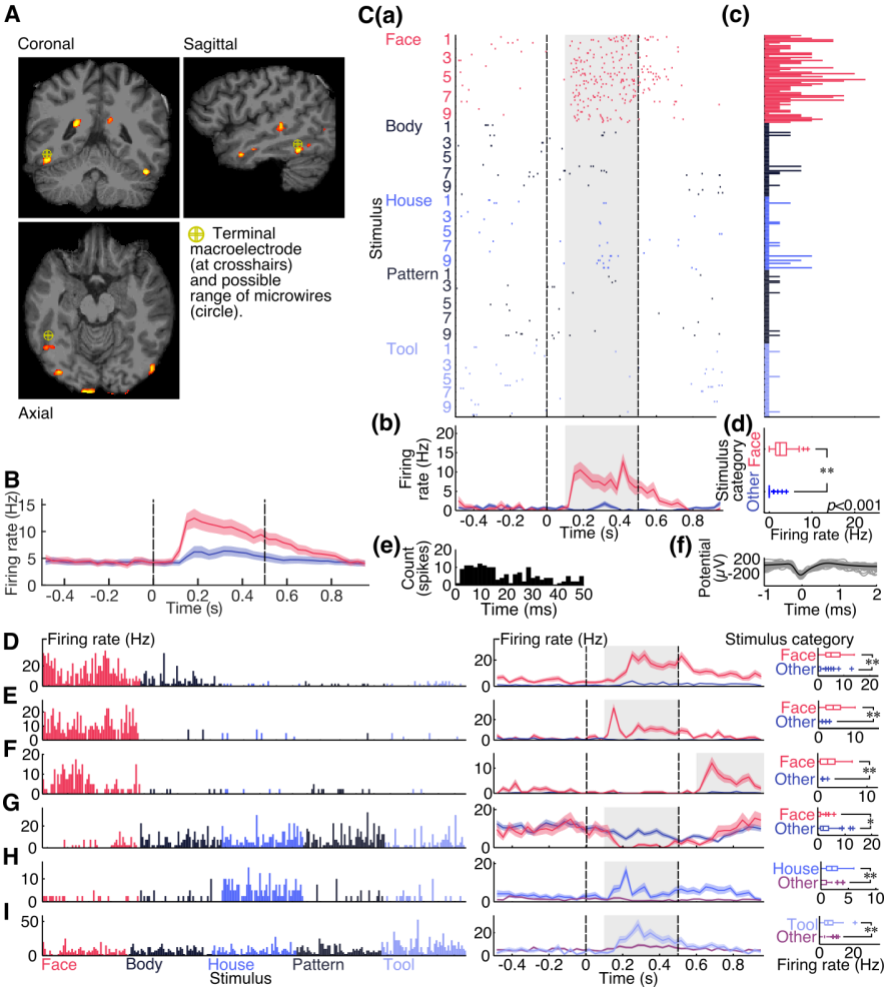

Fig. 1. Diverse category-selective units found in the ventral temporal cortex. (A) Example electrode location (Subject 8) over functional MRI face>house activation contrast map (in red). (B) Grand mean peristimulus-time histogram. Red: faces, blue: non-face objects. Colored shaded areas represent mean ± standard error of the mean (SEM) at each timepoint. Dashed lines represent image onset at time 0 and offset, respectively. (C) Representative face-selective unit from right FFA in Subject 8. (C(a)) Raster plot of responses with each row representing a single trial, and dashed lines representing onset and offset respectively. (C(b)) Mean peristimulus time histogram. Red: faces, blue: non-face objects. Colored shaded areas represent mean ± SEM at each timepoint. (C(c)) Mean firing rate per trial. Average over gray shaded area in raster plot, 0.1 to 0.5 s after presentation. (C(d)) Distribution of responses to faces and non-face objects. Responses to faces are significantly stronger. (C(e)) Inter-spike interval distribution. (C(f)) Spike waveforms. Mean spike waveform in black. (D-I) Peristimulus time histograms and grand mean peristimulus time histograms for units with (D) longer response latency, (E) transient peaked response, (F) offset response, (G) suppression to faces, (H) house selectivity, (I) tool selectivity. *: $p<0.05$; **: $p<0.001$, rank-sum test, Bonferroni-Holm correction.

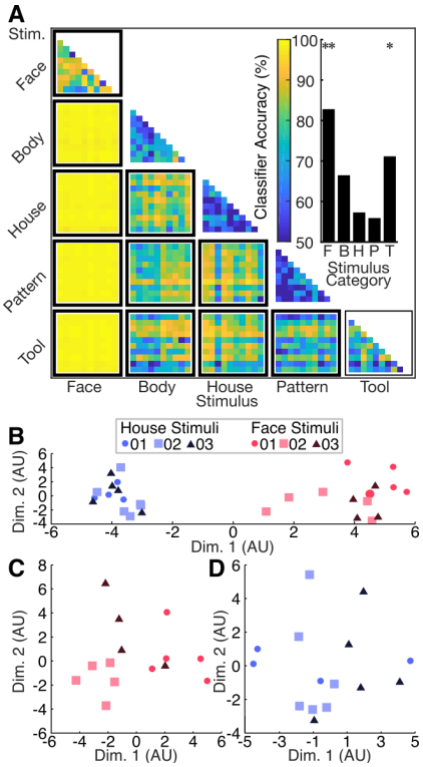

Fig. 2. Exemplar decoding of face stimuli. (A) Binary classifier performance. Cross-validation accuracy of a set of classifiers tested and trained on each pair of exemplars. Discriminability and classifier accuracy of any exemplar pair represented by the color at that location in the matrix, with greater accuracy corresponding to more dissimilar population responses. Inset: Mean decoding accuracies for each intra-category block (classifier accuracy discriminating exemplars of the same category). */thin outline: $p<0.05$; **/thick outline: $p<0.01$, bootstrap 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.

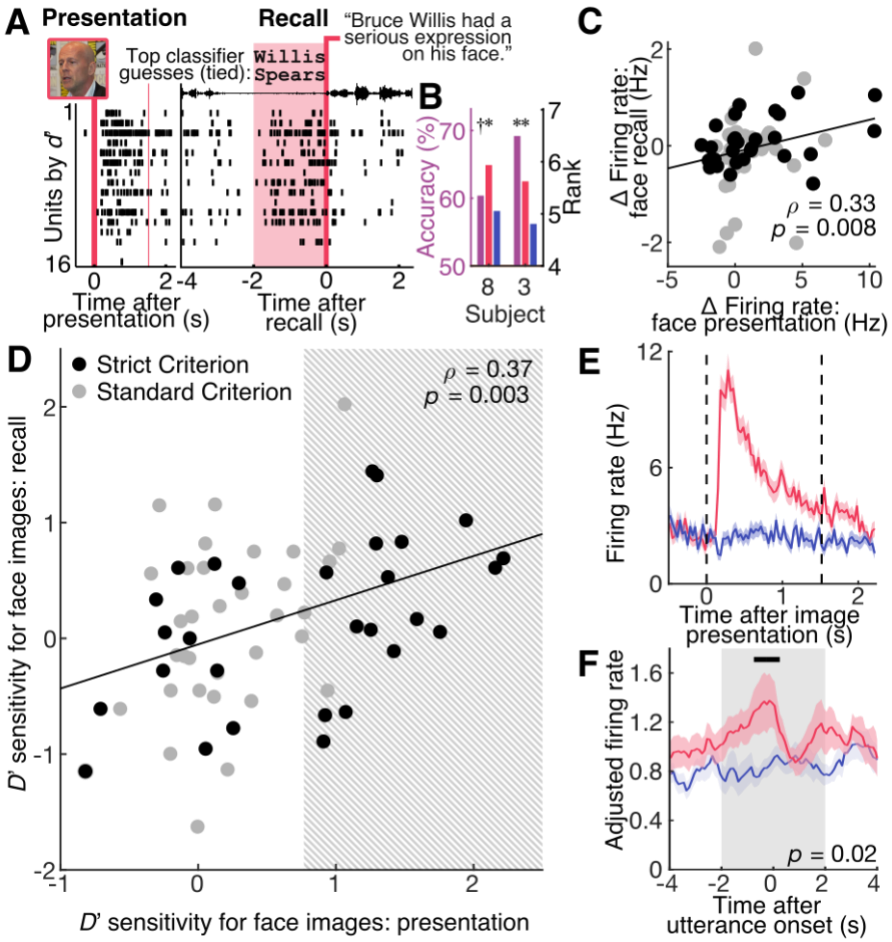

Fig. 3. Face-selective units reactivate during recall. (A) Raster plot from presentation and recall example face stimulus from Subject 8. Image presentation at time 0 on left, and vocalization onset (black envelope trace at top) on the right. Classifier tested on 2 s of activity before recall (red shaded area), predicts that either Bruce Willis (*33*) or Britney Spears will be recalled of seven possible face stimuli. (B) Exemplar decoding accuracy. Cross-category classification: purple bars (left axis). Face exemplars: red bars; places: blue bars, right axis. * $p < 0.05$, bootstrap test, Bonferroni-Holm correction; †: uncorrected. (C) Responses to faces during presentation and before recall. Black: units passing stricter visual responsiveness criterion. (D) Selectivity for faces during presentation and recall. Trend line fitted to all units. (E) Mean peri-stimulus time histogram for face-selective units. Face selectivity defined by gray area in (D). Responses to faces in red and places in blue. Colored areas: mean ± standard error of the mean at each timepoint. Dashed lines represent image onset at time 0 and offset. (F) Mean peri-recall time histogram. Significant difference between face and place activity (gray box, -2 to 2 seconds, 2-sample *t*-test, $p < 0.05$). Black bar: 2-sample *t*-tests, $p<0.05$, uncorrected.