A fast link between face perception and memory in the temporal pole

Sofia M. Landi[1,2,*], **Pooja Viswanathan**[1,3], **Stephen Serene**[1], **Winrich A. Freiwald**[1,4,*]

[1]The Rockefeller University; New York, NY, USA

[2]Department of Physiology and Biophysics, University of Washington, Seattle, WA, USA

[3]The Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[4]The Center for Brains, Minds & Machines, Cambridge, MA, USA.

Abstract

The question of how the brain recognizes the faces of familiar individuals has been important throughout the history of neuroscience. Cells linking visual processing to person memory have been proposed, but not found. Here we report the discovery of such cells through recordings from an fMRI-identified area in the macaque temporal pole. These cells responded to faces when they were personally familiar. They responded non-linearly to step-wise changes in face visibility and detail, and holistically to face parts, reflecting key signatures of familiar face recognition. They discriminated between familiar identities, as fast as a general face identity area. The discovery of these cells establishes a new pathway for the fast recognition of familiar individuals.

One Sentence Summary:

The temporal pole region contains cells linking face perception to person memory.

Recognizing someone we know requires the combination of sensory perception and long-term memory. Where the brain stores these memories, and how it links sensory activity patterns to them, remains largely unknown. Consider the case of person recognition: the same person's face can evoke vastly different retinal activity patterns, yet all activate the

---

Supplementary Materials:

Materials and Methods

Figs. S1 to S6

References (39–46)

same person's memory. We know how information from the eyes is transformed to extract facial identity across varying viewing conditions in the face processing network (1), but not where and how this representation then activates person memory.

Theories for the neural basis of person recognition have a long history in neuroscience dating back to the idea of the "grandmother neuron" in the 1960s, which would respond to any image of one's grandmother and support the recollection of grandmother-related memories (2). A later theory posited a hybrid "face recognition unit" (3), which would combine properties of sensory face cells in encoding facial information with properties of memory cells in storing information from past personal encounters. Yet neither class of neuron has been found.

Face cells and an entire network of face areas have been discovered in the superior temporal sulcus (STS) and inferotemporal (IT) cortex (1, 4, 5), and person memory cells in the medial temporal lobe (6). However, in the temporal pole only few electrophysiological recordings have been performed (7). With neuropsychological evidence pointing towards a role of this region in person recognition (8), and the recent discovery of a small sub-region (TP) selective for familiar faces (9), we decided to record from the temporal pole. Because face identity memories might be consolidated exactly where they are processed (10), we also recorded from the most identity-selective face area in IT, the anterior-medial face area (AM) (1, 11) (Fig. 1A).

Using whole-brain functional magnetic resonance imaging (fMRI) we localized areas TP and AM in the right hemispheres of two rhesus monkeys (Figs. 1B, 1C and S1A, see Methods). We recorded responses from all cells encountered. We assessed visual responsiveness, visual selectivity, and familiarity selectivity with a 205-image set that included human faces (30 personally familiar, 30 unfamiliar), monkey faces (12 personally familiar, 1 subject's own face, 72 unfamiliar), bodies (15 unfamiliar), objects (15 personally familiar, 25 unfamiliar), and gray background (5 images).

An example cell from area TP (Fig. 1D left) remained visually unresponsive to any of the 145 face stimuli – with one exception (stimulus 33), the face of a personally familiar monkey. Another example cell (Fig. 1D right) was unresponsive to non-face stimuli and responded selectively to the faces of several familiar monkeys. This pattern of high visual responsiveness, preference for monkey faces, and selectivity for familiar faces was typical for the TP population as a whole (Fig. 2A, Fig S2): 90 out of 98 (92%) neurons responded significantly to at least one image (see Methods), and the TP population preferred monkey over human and familiar over unfamiliar monkey faces (Figs. 2A & 2C left, significant 2-way ANOVA stimulus category x familiarity: $F(2,18124) = 89.61$, $p < 10^{-39}$, post-hoc Tukey's HSD test: $p < 10^{-4}$). Selectivity for face familiarity was so high, the TP population responded more than three times as much to familiar than to unfamiliar monkey faces. AM cells, though also visually responsive (125 of 130, 96%), responded similarly to both monkey and human and familiar and unfamiliar faces (Figs. S1B, 2B, 2C & S2, interaction effects $F(2,24044) = 2.22$, $p > 0.1$, familiarity effect $F(1,24044) = 0.3$, $p>0.1$, stimulus category $F(2,24044) = 317.99$, $p < 10^{-100}$). The effect of familiarity on neuronal responses differed between TP and AM and between stimulus categories (three-way

ANOVA interaction effect Area x Familiarity x Category, F(2,42168) = 31.29, p<0.001). While in AM there were no significant familiarity effects for any category (p>0.1), in TP familiar monkey faces elicited a significantly higher response than all other categories (post hoc tests, $p<10^{-4}$ corrected using Tukey's HSD).

The pattern of the TP population response was also specialized for familiar faces: population responses were most similar for familiar monkey faces (Fig. 2D left) and were adjacent and separate from unfamiliar monkey faces in a 2D representational space (Fig. 2E left), supporting accurate decoding of only the familiar monkey category (Fig. 2F left, see Methods). In AM, population response similarity was high for all categories, and stimuli belonging to the same category, whether familiar or not, clustered together (Fig. 2D & 2E right). While the separability of faces and objects was higher for AM (separability index SI = 0.57±0.01) than TP (SI = 0.26±0.01, permutation test p<0.005), the separability of a monkey familiar face cluster was higher in TP (SI = 0.73±0.02, see Methods) than in AM (SI = 0.43±0.03, permutation test p<0.005). This fundamental difference between TP and AM was also reflected in category decoding results (Fig. 2F). Crucially, TP's familiarity selectivity did not result from passive visual exposure – subjects saw all pictures thousands of times– but rather from real-life personal encounters.

TP cells express one key property of face recognition units (3): modulation by face familiarity. To achieve recognition, face recognition units also need to discriminate facial identities (as the TP example cells in Fig. 1C). Population decoding analyses within each of the four face categories (see Methods) showed that the TP population discriminated between identities of familiar monkey faces, and only between these (Fig. 3A & S3A, left panel, p < 0.005 permutation test). In contrast, AM reliably discriminated between identities in each face category (Fig. 3A & S3A right panel, p < 0.005 permutation tests). TP and AM encoded the identity of familiar monkeys through a mix of sparse and broadly tuned cells. The mean sparseness indexes for familiar monkey faces (see Methods) was 0.65 in TP and 0.61 in AM (p = 0.26, Wilcoxon test), ranging in both areas from approximately 0.08 to 0.97.

Next, we tested whether TP cells show three functional signatures resembling the psychophysics of human face recognition. First, we determined whether TP responses exhibit an all-or-none perceptual threshold, as face detection does (12, 13). We created visual stimuli with different levels of phase scrambling for a given cell's preferred familiar and unfamiliar faces (Fig. 3B, top). Both TP and AM cells responded from a specific visibility threshold on (Figs. 3B). We fit the spiking response of each cell to a sigmoidal function (see Methods), whose exponent quantifies the steepness of the non-linear effect. Response steepness depended on the interaction of stimulus familiarity and area (two-way ANOVA interaction effect, F(1,100) = 6.89, p<0.01). Post hoc tests revealed no significant differences in non-linearity between familiar and unfamiliar faces in AM (p=0.8) but a higher non-linearity for familiar faces in TP (p<0.01).

Second, we tested whether TP can directly and specifically support the recognition of familiar faces. In another experiment, we applied 10 steps of Gaussian blurring (14) to the preferred familiar and unfamiliar faces of each cell (Fig. 3C, top). Quantifying non-linearity as above, we found response steepness to depend on the interaction of stimulus familiarity

and area (Fig. 3C, two-way ANOVA interaction effect, $F(1,100) = 4.16$, p<0.05). In TP only familiar faces elicited this non-linear response, while AM cells failed to show this effect altogether (Fig. S3B, 3C, post hoc tests corrected using Tukey's HSD, p<0.05).

Finally, another fundamental psychophysical property of familiar face recognition is that internal facial features are most informative about a familiar identity (15–17). We cropped familiar faces into an inner and outer component and then the inner one further into eyes, mouth, and nose (Fig. 3D, top). These manipulations significantly impacted TP (n = 27, one-way ANOVA yielded a significant effect of stimulus type on the spiking response [p < 0.001, $F(5,157) = 48.07$]). TP cells responded almost as much to the inner face alone as to the whole face (post hoc tests corrected using Tukey's HSD, p = 0.99), but only weakly to the outer face alone (Tukey's HSD, p < 0.01) (Fig. 3D & S3C, left). Responses to the isolated inner face parts were minimal and did not add up to the response of the inner face. We found a similar non-linear effect in TP, but not AM, when using decomposed face images in different frequency bands (Fig. S3D).

TP neurons are highly face selective. However, they might respond not only to the face, but to the whole social agent (i.e. face + body), thus resembling person identity nodes (3). However, TP's population responses to body images were weaker than to face images (Fig 1C, two-sample Kolmogorov-Smirnov test, p<0.01, compare Figs. S4A and B). To test whether body context could augment TP responses to faces (18), we presented pictures of entire familiar individuals, their isolated faces, and their isolated bodies (Fig. S4C). In 27 TP cells tested, the response to the face alone was almost as strong as that to the entire individual (0.8 spikes/s difference at response peak, permutation tests, p < 0.05, Fig. S4D), while isolated familiar bodies did not elicit a clear response.

Human psychophysics has found familiar face recognition to be robust to identity-preserving transformations (3), but suggested different encoding schemes (e.g. (19)). We tested tuning to two identity-preserving transformations: (1) in-depth head rotation, a transformation that has been characterized in face areas including AM before (1); and (2) geometric image distortion that does not affect familiar face recognition (20). We found that TP cells are as robust to in-depth rotation as AM cells: a three-way ANOVA with in-depth rotation, face identity and area as factors yielded no significant interaction effects between any of the factors (Fig. S5A–D View x Identity x Area $F(48,2535)=0.201$, p>0.9; Area x Identity $F(12,2532) = 0.63$, p>0.8; Area x View $F(4,2535)=0.22$, p>0.9; Identity x View, $F(48,2532)=0.34$, p>0.9). Geometric deformations had no distinctive effect on AM and TP population responses (Fig. S5E–F, two-way ANOVA with distortion type and area as factors, interaction effects $F(6,350)=0.34$, p>0.9).

The dominant models of face recognition (3, 21, 22) posit a sequential transition from perceptual face identity processing to face or person recognition. AM is located at the pinnacle of perceptual face processing (1), exhibiting an efficient code for physical face identity (11, 23). If face recognition units in TP are downstream from AM, their response latencies should be systematically longer. We tested this prediction in three analyses.

First, population response latencies of the Face and Object Familiarity experiment (Fig. 1D, S1B, Figs. 4A&B), were not systematically different between face categories or areas (Figs. 4A&B, Fig. S6, no significant interaction [F(3,312) = 1.17, p > 0.3] or main effects for category [F(3,312) = 0.93, p > 0.4] and brain area [F(1,312) = 3.3, p > 0.05]).

Second, we analyzed performance time courses for five binary decoders in TP and AM (Fig. 4C, see Methods): faces vs. non-faces (Face detection), human vs. monkey faces (Species identification), and familiar vs. unfamiliar monkey faces, human faces, and objects (Semantic classification). Peak decoding accuracy for face detection and species identification was higher for AM than TP (p<0.05, permutation tests). The reverse was the case for semantic classification of monkey faces (p<0.05, permutation tests, Fig. 4D), which seems to emerge faster in TP. However, we couldn't detect any significant latency differences between AM and TP for any of the binary classifiers (p>0.1, permutation tests, Fig. 4D).

Third and last, we analyzed the time courses of identity decoding between familiar monkey faces in TP and AM (Figs. 4E–F). Onset and peak times for decoding familiar monkey identities were similar in both areas (p>0.1, permutation tests, Fig. 4F).

We report here the discovery of a new class of face memory cells. They share with Lettvin's grandmother cell hypothesis (2), Konorski's "gnostic unit" (24), and Bruce & Young's face recognition unit (3) the conjunction of facial shape and familiarity selectivity. TP differs, however, from the grandmother cell hypothesis in that the identity of a familiar face is not represented by a single neuron, but by a distributed population response. While highly face selective, TP cells are qualitatively different from inferotemporal face cells, even those at the apex of the face-processing system in area AM (1, 11): the new cells are selective not for faces in general, but for personally familiar ones; encode personally familiar faces both categorically and individually; and exhibit key functional characteristics of face recognition. They also differ from mediotemporal person concept cells (6) by a much shorter response latency and selectivity towards the inner face. TP cells encode familiar face identities not by single cells - as the grandmother neuron concept (2) suggested - but as populations (24).

Past studies have found visual familiarity with a stimulus to reduce activity throughout object and face recognition systems, and described this reduction as repetition suppression, predictive normalization, or sparsification (9, 25–29). These effects result primarily from repeated stimulus exposure. Our finding of (i) selective and specific response enhancement (ii) that is robust across multiple transformations (iii) in a spatially localized brain region (iv) outside of core object and face processing systems (v) as a result of personal real-life experience, is a fundamentally different memory mechanism. Memory consolidation theories agree that long term memories are stored in the cortex (10, 30). Here we show that personal real-life experience has the astonishing capacity to carve out a small piece of cortex, and consolidate very specific memories there. If familiar conspecific face memories are stored in one small region of the temporal pole, other modules with similar specificity probably exist nearby. More complex knowledge systems, e.g. about individuals and their social relationships (31) may be built upon these foundations. This would explain person-related agnosia following damage to the temporal pole (8).

TP signals face information surprisingly fast, which might explain the astonishing speed of familiar face recognition (32, 33). The simultaneity of familiar face processing in TP and AM and the qualitative differences in their selectivity – TP functionally mimics face recognition, while AM does not – suggest that AM and TP may operate functionally and possibly structurally in parallel. For example, a specific subset of short-latency AM cells may provide face-identity information to TP. Alternatively, in agreement with the lack of documented direct connections between AM and TP (34), there may be two pathways of face and person memory: one pathway from AM to a perirhinal face area (9), entorhinal cortex, and the hippocampus, and a second pathway to TP. The first pathway would facilitate the formation of new associations (35–37) and the feeling of familiarity (38). The second pathway would allow for direct access - without the need to recapitulate all stages of the first pathway - to long-term semantic face information in the temporal pole.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

Data and materials availability:

There are no restrictions on data availability and data will be deposited on FigShare repository (https://www.figshare.com) under accession number http://www.dx.doi.org/10.6084/m9.figshare.14642619.

References and Notes

1. Freiwald WA, Tsao DY, Science. 330, 845–851 (2010). [PubMed: 21051642]

2. Gross CG, Neuroscientist. 8, 512–518 (2002). [PubMed: 12374433]

3. Bruce V, Young A, Br. J. Psychol. 77, 305–327 (1986). [PubMed: 3756376]

4. Perrett DI, Hietanen JK, Oram MW, Benson PJ, Philos. Trans. R. Soc. Lond. B Biol. Sci. 335, 23–30 (1992). [PubMed: 1348133]

5. Kanwisher N, McDermott J, Chun MM, The Journal of neuroscience : the official journal of the Society for Neuroscience. 17, 4302–4311 (1997). [PubMed: 9151747]

6. Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I, Nature. 435, 1102–1107 (2005). [PubMed: 15973409]

7. Nakamura K, Matsumoto K, Mikami A, Kubota K, Journal of Neurophysiology. 71, 1206–1221 (1994). [PubMed: 8201413]

8. Olson IR, Plotzker A, Ezzyat Y, Brain. 130, 1718–1731 (2007). [PubMed: 17392317]

9. Landi SM, Freiwald WA, Science. 357, 591–595 (2017). [PubMed: 28798130]

10. Barry DN, Maguire EA, Trends Cogn Sci (Regul Ed). 23, 635–636 (2019).

11. Chang L, Tsao DY, Cell. 169, 1013–1028.e14 (2017). [PubMed: 28575666]

12. Gobbini MI et al., PLoS ONE. 8, e66620 (2013). [PubMed: 23805248]

13. Sinha P, Balas B, Ostrovsky Y, Russell R, Proc. IEEE. 94, 1948–1962 (2006).

14. Ramon M, Vizioli L, Liu-Shuang J, Rossion B, Proc Natl Acad Sci USA. 112, E4835–44 (2015). [PubMed: 26283361]

15. Andrews TJ, Davies-Thompson J, Kingstone A, Young AW, J. Neurosci. 30, 3544–3552 (2010). [PubMed: 20203214]

16. Ellis HD, Shepherd JW, Davies GM, Perception. 8, 431–439 (1979). [PubMed: 503774]

17. Young AW, Hay DC, McWeeny KH, Flude BM, Ellis AW, Perception. 14, 737–746 (1985). [PubMed: 3837875]

18. Fisher C, Freiwald WA, Proc Natl Acad Sci USA. 112, 14717–14722 (2015). [PubMed: 26464511]

19. Burton AM, Jenkins R, Hancock PJB, White D, Cogn. Psychol. 51, 256–284 (2005). [PubMed: 16198327]

20. Sandford A, Burton AM, Cognition. 132, 262–268 (2014). [PubMed: 24853629]

21. Gobbini MI, Haxby JV, Neuropsychologia. 45, 32–41 (2007). [PubMed: 16797608]

22. Natu V, O'Toole AJ, Br. J. Psychol. 102, 726–747 (2011). [PubMed: 21988381]

23. Koyano KW et al., Curr. Biol. 31, 1–12.e5 (2021). [PubMed: 33065012]

24. Konorski J, Chicago: University of Chicago Press. (1967).

25. Xiang JZ, Brown MW, Neuropharmacology. 37, 657–676 (1998). [PubMed: 9705004]

26. Freedman DJ, Riesenhuber M, Poggio T, Miller EK, Cereb. Cortex. 16, 1631–1644 (2006). [PubMed: 16400159]

27. Meyer T, Olson CR, Proc Natl Acad Sci USA. 108, 19401–19406 (2011). [PubMed: 22084090]

28. Woloszyn L, Sheinberg DL, Neuron. 74, 193–205 (2012). [PubMed: 22500640]

29. Schwiedrzik CM, Freiwald WA, Neuron. 96, 89–97.e4 (2017). [PubMed: 28957679]

30. Scoville WB, Milner B, J. Neurol. Neurosurg. Psychiatr. 20, 11–21 (1957).

31. Sliwa J, Freiwald WA, Science. 356, 745–749 (2017). [PubMed: 28522533]

32. Dobs K, Isik L, Pantazis D, Kanwisher N, Nat. Commun. 10, 1258 (2019). [PubMed: 30890707]

33. Visconti di Oleggio Castello M., Gobbini MI, PLoS ONE. 10, e0136548 (2015). [PubMed: 26305788]

34. Grimaldi P, Saleem KS, Tsao D, Neuron. 90, 1325–1342 (2016). [PubMed: 27263973]

35. Miyashita Y, Nature. 335, 817–820 (1988). [PubMed: 3185711]

36. Ison MJ, Quian Quiroga R., Fried I, Neuron. 87, 220–230 (2015). [PubMed: 26139375]

37. Miyashita Y, Nat. Rev. Neurosci. 20, 577–592 (2019). [PubMed: 31485007]

38. Tamura K et al., Science. 357, 687–692 (2017). [PubMed: 28818944]

39. Ohayon S, Tsao DY, J. Neurosci. Methods. 204, 389–397 (2012). [PubMed: 22192950]

40. Paxinos G, Huang XF, Toga AW, (2000).

41. Willenbockel V et al., Behav. Res. Methods. 42, 671–684 (2010). [PubMed: 20805589]

42. Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS, Science. 311, 670–674 (2006). [PubMed: 16456083]

43. Rolls ET, Critchley HD, Browning AS, Inoue K, Exp. Brain Res. 170, 74–87 (2006). [PubMed: 16328289]

44. Dehaqani M-RA et al., J. Neurophysiol. 116, 587–601 (2016). [PubMed: 27169503]

45. Meyers EM, Front. Neuroinformatics. 7, 8 (2013).

46. Naka KI, Rushton WA, The Journal of physiology. 185, 536–555 (1966). [PubMed: 5918058]
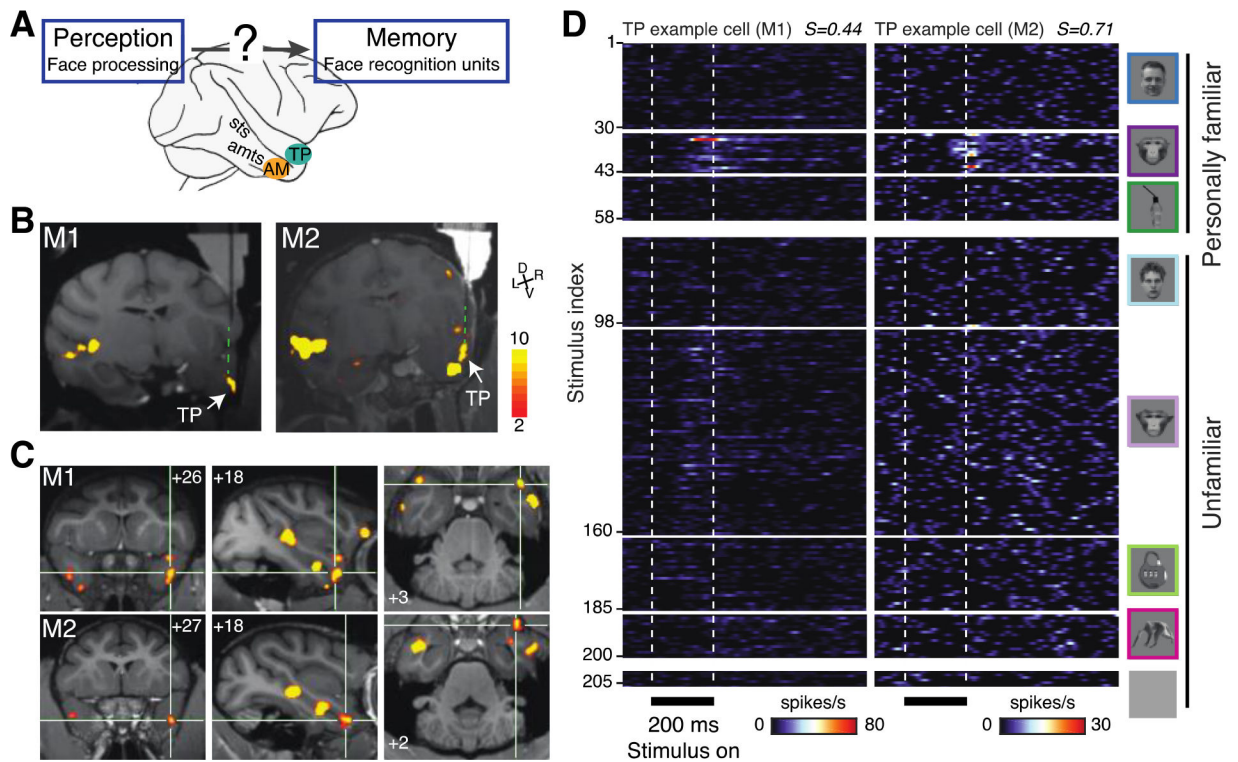
Fig. 1: Cells in the temporal pole area TP respond to familiar faces
(**A**) Schematic: face perception systems are thought to feed into downstream face memory systems (3) in ways yet unknown. Candidate areas in the macaque brain are area AM and area TP (**B**) Structural MRI (T1) and functional overlay (faces>objects, color coded for negative common logarithm of p value (p<0.001, uncorrected) showing electrodes targeting recording TP in monkeys M1 and M2 (see Methods). (**C**) Coronal (left), parasagittal (middle) and axial (right) slices showing TP in M1 and M2. Numbers indicate stereotaxic coordinates: mm rostral to interaural line, mm from midline to the right, and mm dorsal from interaural line, respectively. (**D**) Mean peri-stimulus time histograms of two TP example cells (left M1, right M2) responding to 205 stimuli set (Face Object Familiarity, FOF) in eight categories (top to bottom, far right) presented for 200ms (bottom) with 500ms inter-stimulus intervals in spikes per second (color scale bottom). Each cell responds significantly to a range of familiar monkey faces. Sparseness indexes (see Methods) are shown in the top right of each plot.
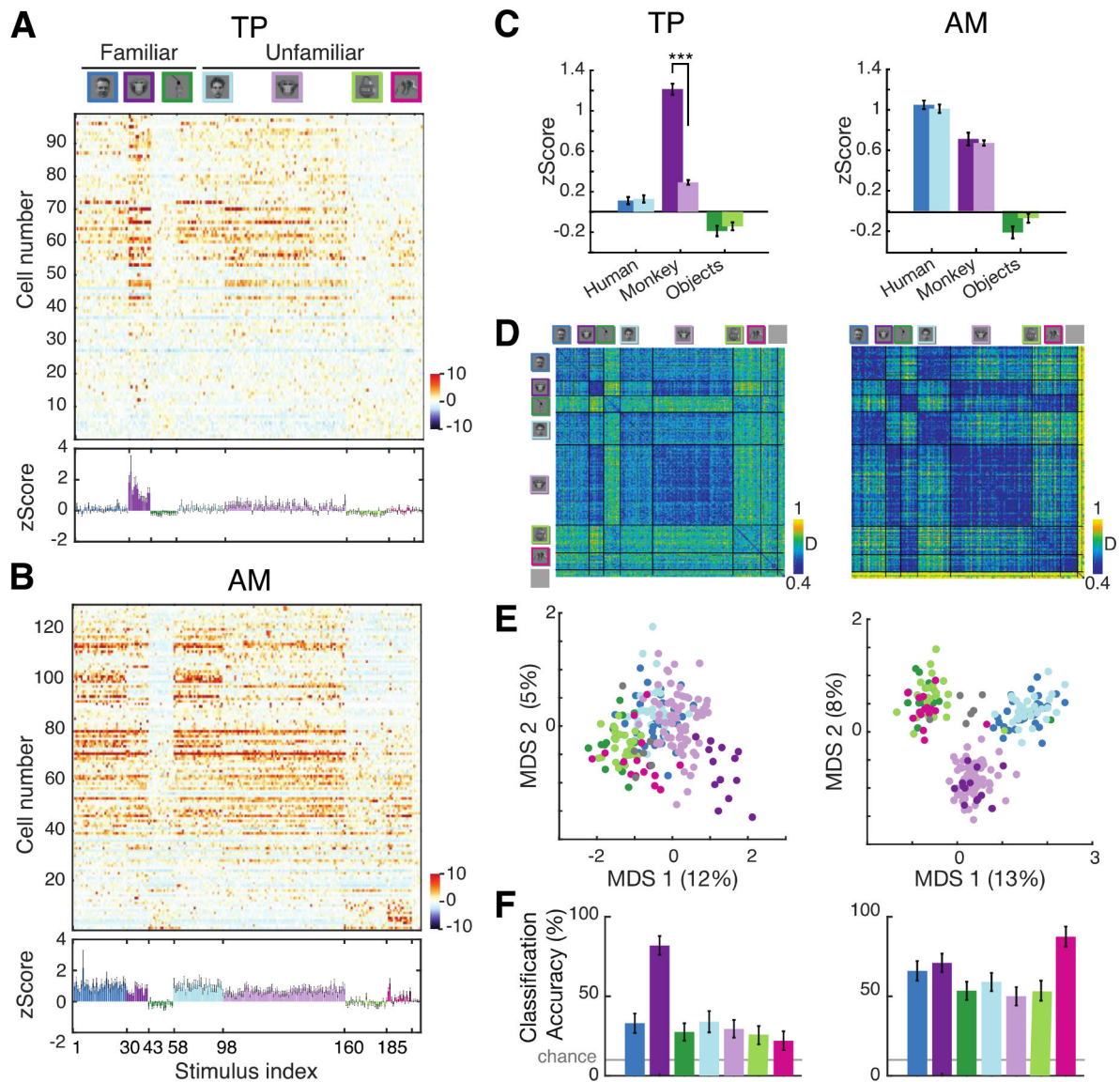
Fig. 2: TP is selective for familiar monkey faces, but not other familiar stimuli.

(**A**) Population response matrices (z-scored, color scale lower right) to FOF stimulus set (top) for all recorded TP cells (n=98, sorted top to bottom by face selectivity index, FSI, see Methods). Bottom: average population response (mean z-Score ± SEM). (**B**) Same as (A) for AM (n=130). (**C**) TP (left) and AM (right) population response (average z-Scores, error bars indicating 95% confidence intervals) for six categories (color scales as in A). Significant post hoc tests (***, p<10⁻⁴, corrected using Tukey's HSD) are shown for familiar versus unfamiliar stimuli. (**D**) TP (left) and AM (right) population response dissimilarity matrix (1-Pearson correlation coefficient) for FOF stimulus set, color scale lower right. (**E**) Individual stimuli in two-dimensional space derived from MDS of dissimilarity. The explained variance is shown for each dimension in the axis labels. (**F**) TP (left) and AM (right) population category decoding performance measured by linear classifier performance (see Methods).
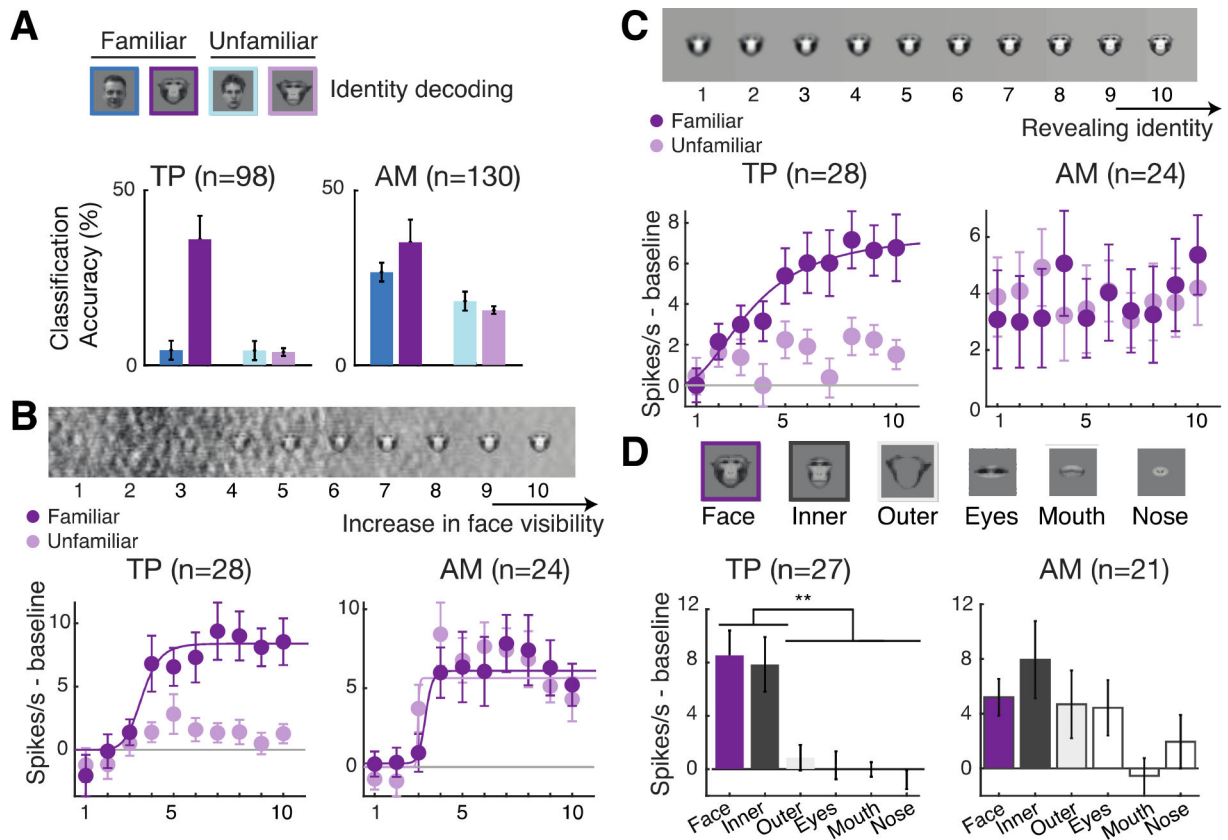
Fig. 3. TP encodes identity information and mimics psychophysics of face recognition.
(**A**) Decoding accuracies (% normalized classification accuracy) in TP (left) and AM (right) for identity discrimination among familiar human faces (dark blue), familiar monkey faces (violet), unfamiliar human faces (light blue), and unfamiliar monkey faces (lilac). (**B**) TP and AM population responses to pictures of familiar and unfamiliar faces at ten levels of phase scrambling (100%−0% left to right). Mean responses, SEM (error bars), and significant sigmoidal functions fits (solid lines, only shown for conditions with $R^2 > 0.5$) are shown. (**C**) TP and AM population responses to pictures of familiar and unfamiliar faces at ten blurring levels (see Methods). Populations and conventions as in B. (**D**) TP and AM population responses to pictures of whole and cropped faces (top). Populations and conventions as in B and C. Significant post hoc tests (**, $p < 0.01$, corrected using Tukey's HSD).
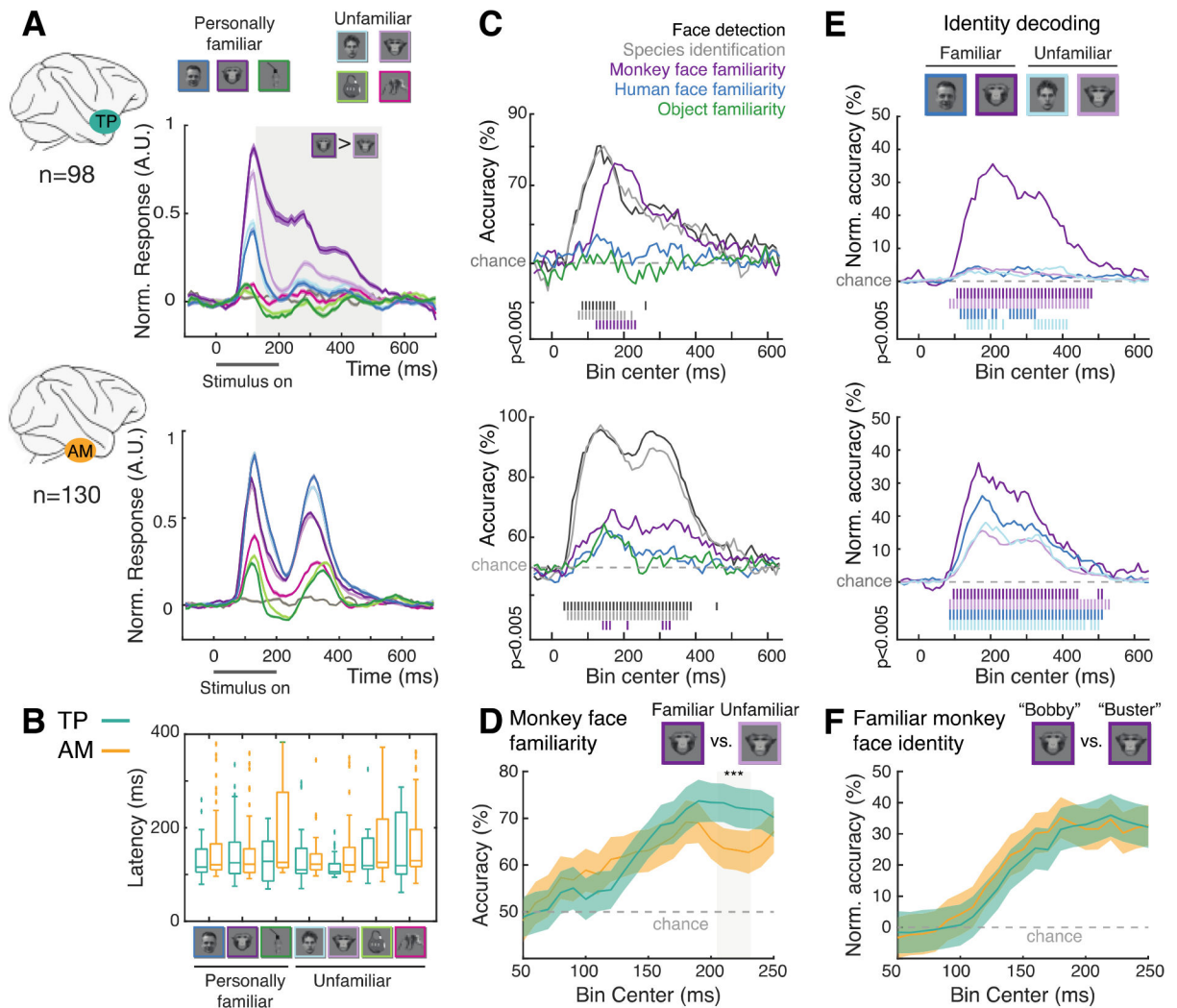
Fig. 4. Simultaneous and early familiar face processing in TP and AM.
(**A**) Average normalized PSTHs for TP (top) and AM (bottom)for all categories. Color shading indicates SEM, gray shaded area time period of significantly larger responses to familiar than unfamiliar monkey faces (permutation tests, 1000 iterations, p<0.01). Color code shown in the top of the figure, the gray line represents the average response to gray background images (no visual stimuli). (**B**) Response latencies in TP (green) and AM (yellow) for all categories. (**C**) Time courses of decoding performance of TP (top) and AM (bottom) population response for five contrasts. Vertical bars below plot indicate significant decoding accuracies (permutation tests, n = 200, p < 0.005, see Methods). (**D**) Fine-scale peri-stimulus time course of monkey familiarity information in TP (green) and AM (yellow). Error bars are the SD of the decoding accuracies over all the shuffled trials (repeated n=200) and cross-validation splits. (**E**) Time courses of decoding performance of TP (top) and AM (bottom) population response for within-category (top) face identification (% normalized classification accuracy). Populations and conventions as in C. (**F**) Fine-scale peri-stimulus

time course of familiar monkey identity information in TP (green) and AM (yellow). Error bars were calculated as in (D).