RESEARCH ARTICLE

NEUROSCIENCE

Cross-modal representation of identity in the primate hippocampus

Timothy J. Tyree[1,2], Michael Metke[1,3], Cory T. Miller[1,3]*

Faces and voices are the dominant social signals used to recognize individuals among primates. Yet, it is not known how these signals are integrated into a cross-modal representation of individual identity in the primate brain. We discovered that, although single neurons in the marmoset hippocampus exhibited selective responses when presented with the face or voice of a specific individual, a parallel mechanism for representing the cross-modal identities for multiple individuals was evident within single neurons and at the population level. Manifold projections likewise showed the separability of individuals as well as clustering for others' families, which suggests that multiple learned social categories are encoded as related dimensions of identity in the hippocampus. Neural representations of identity in the hippocampus are thus both modality independent and reflect the primate social network.

Navigating complex primate societies relies on learning the identity of each individual in the group and their respective social relationships (*1*). Neurons in the brains of primates and other mammals selectively respond to the identity when viewing the face or hearing the voice of a specific individual as unimodal signals (*2*–*8*). However, data showing that single neurons are responsive to both the face and voice of an individual—a cross-modal representation of identity—are limited to "concept cells" in the human hippocampus (*9*–*11*). These neurons are notable for several reasons, including their putative role in memory functions (*12*) and potential uniqueness to humans (*13*). We investigated whether cross-modal representations of identity are evident in the hippocampus of marmoset monkeys by recording single hippocampal neurons (*14*) while presenting subjects with multiple exemplars of individual marmoset faces (from different viewpoints) and voices as unimodal stimuli (*4*), as well as concurrently by presenting the faces and voices from the same or different individuals—i.e., match versus mismatch (MvMM). Visual stimuli were presented from a monitor directly in front of the animal while a speaker positioned directly below the screen broadcast the acoustic stimuli. Subjects were only presented with familiar conspecifics housed in the same colony who differed in their respective social relatedness (*11*).

Identity-selective neurons

To first test whether cross-modal representations of identity are evident in the hippocampus of a nonhuman primate, we performed the same receiver operator characteristic (ROC) selectivity analysis described previously in humans (*9*–*11*) and detected a population of cross-modal invariant neurons for individual identity when observing marmoset faces or voices (Fig. 1A and fig. S1A), as well as neurons selective for individual identity when viewing only their faces (Fig. 1B and fig. S1B) or hearing only their voices (Fig. 1C and fig. S1C). These identity neurons were confirmed in all hippocampal subfields (Fig. 1D). Notably, only neurons that exhibited a mean peak firing rate 2 SDs above baseline qualified for this analysis, which supports $P < 0.01$, and differed from the 5 SDs used previously in humans (*9*–*11*). Responses were determined from the mean firing rate during a 500-ms continuous sliding window maximized over the duration of the 3500-ms stimulus. Overall, we observed that $N = 148$ (9.2%) of $N = 1602$ qualifying neurons demonstrated selectivity for a single preferred individual (Fig. 1E), with different neurons selective for faces ($N = 52$), voices ($N = 39$), or both faces and voices ($N = 57$) (Fig. 1F). The mean area under the ROC curve (AUC) of identity neurons (AUC = 0.902 ± 0.014) was significantly above chance ($P < 0.001$) (Fig. 1G). Although these neurons in marmosets were overall less selective than in humans (*9*–*11*), this disparity may reflect species differences in hippocampus properties that affect neural coding mechanisms for identity. Baseline hippocampal activity, for example, was considerably higher in the current study (mean 6.47 Hz; $N = 2358$ neurons) (fig. S2) than has been reported in humans (*15*), although a more comprehensive comparative analysis of physiological differences is needed to better understand how such differences affect hippocampal functions.

Analysis of eye movements (Fig. 1H) revealed that marmosets' visual behavior and neural activity were differentially affected by mo[illegible] ity and identity. Marmosets exhibited sig[illegible] cantly shorter fixations ($P < 0.001$, $N_{fixations} = 18{,}965$) (Fig. 1I) and significantly more saccades ($P < 0.001$, $N_{saccades} = 2203$) during trials with face-only relative to the voice-only trials (Fig. 1J). These monkeys were also highly focused on faces during stimulus presentations, with faces accounting for 77.9% of viewing time and eyes specifically accounting for 37.6% of viewing time. The firing rate of identity neurons was significantly greater than the remaining neurons when subjects were looking at the eyes or face (both $P < 0.001$) (Fig. 1K). This was not, however, a broad attentional effect (*16*) because the firing rate of simultaneously recorded nonidentity neurons did not show the same increased firing rate when gazing at faces or eyes.

Multiple identities are represented in single neurons

A potential parallel mechanism to highly selective concept cells is for individual cells to contribute to multiple functions (*17*, *18*), such as single neurons being sensitive to the cross-modal identity of multiple conspecifics. Hippocampal neurons are sensitive to mismatches between the features of a particular stimulus and a previously learned category (*19*, *20*). To test whether a similar mechanism is evident for the learned social identities of conspecifics in marmoset hippocampus, we tested whether neurons would respond differently when simultaneously observing the face and voice from the same (identity match) or different (identity mismatch) individuals. By presenting a face and voice in all identity MvMM trials, we controlled for the potential effects of multimodal integration (fig. S3A) and instead tested whether a subordinate category, identity, elicited changes in neural activity. Indeed, a subpopulation of units, MvMM neurons, exhibited a significant firing rate preference for either match trials (Fig. 2A) or mismatch trials (Fig. 2B), with some neurons modulated only by this category distinction (Fig. 2A) and others more generally stimulus driven (Fig. 2B). Overall, 21.7% of neurons ($N = 511$ of 2358) exhibited a significant response during MvMM trials, with significantly more units exhibiting a higher firing rate during match ($N = 401$) than mismatch ($N = 110$) trials ($P < 0.001$) (Fig. 2C and fig. S3B). MvMM neurons were largely distinct from the identity neurons described above (Fig. 2D and fig. S3C). Notably, 56% of the neurons observed in both populations whose anatomical location could be confirmed were recorded in CA1. In contrast to identity neurons, MvMM neurons were biased to CA1 (Fig. 2E), with $N = 155$ (44.3%) out of 350 neurons confirmed in the CA1 qualifying as MvMM neurons. In CA1, significantly more MvMM neurons ($N = 129/155$, 83.2%) preferred match

[1]Cortical Systems and Behavior Laboratory, University of California San Diego, La Jolla, CA 92039, USA. [2]Department of Physics, University of California San Diego, La Jolla, CA 92039, USA. [3]Neurosciences Graduate Program, University of California San Diego, La Jolla, CA 92039, USA.
*Corresponding author. Email: corymiller@ucsd.edu

trials to mismatch trials ($P < 0.001$). MvMM neurons exhibited significantly higher median firing rate while the subject was looking at the eyes or face ($P < 0.001$, $N = 511$) (Fig. 2F). Marmosets exhibited significantly more saccadic eye movements during mismatch trials (Fig. 2G), and this difference in behavior was most prominent 1 to 2 s after stimulus onset ($P < 0.05$, $N_{saccades} = 4603$) (Fig. 2H).

Encoding cross-modal identity in neuron populations

These findings suggest that two seemingly distinct mechanisms for representing cross-modal identity are evident in primate hippocampus. We conjectured that more temporally selective coding mechanisms in hippocampus may inform how these two processes for encoding identity are integrated at a population level. Therefore, we developed an algorithm to identify intervals of time during which individual

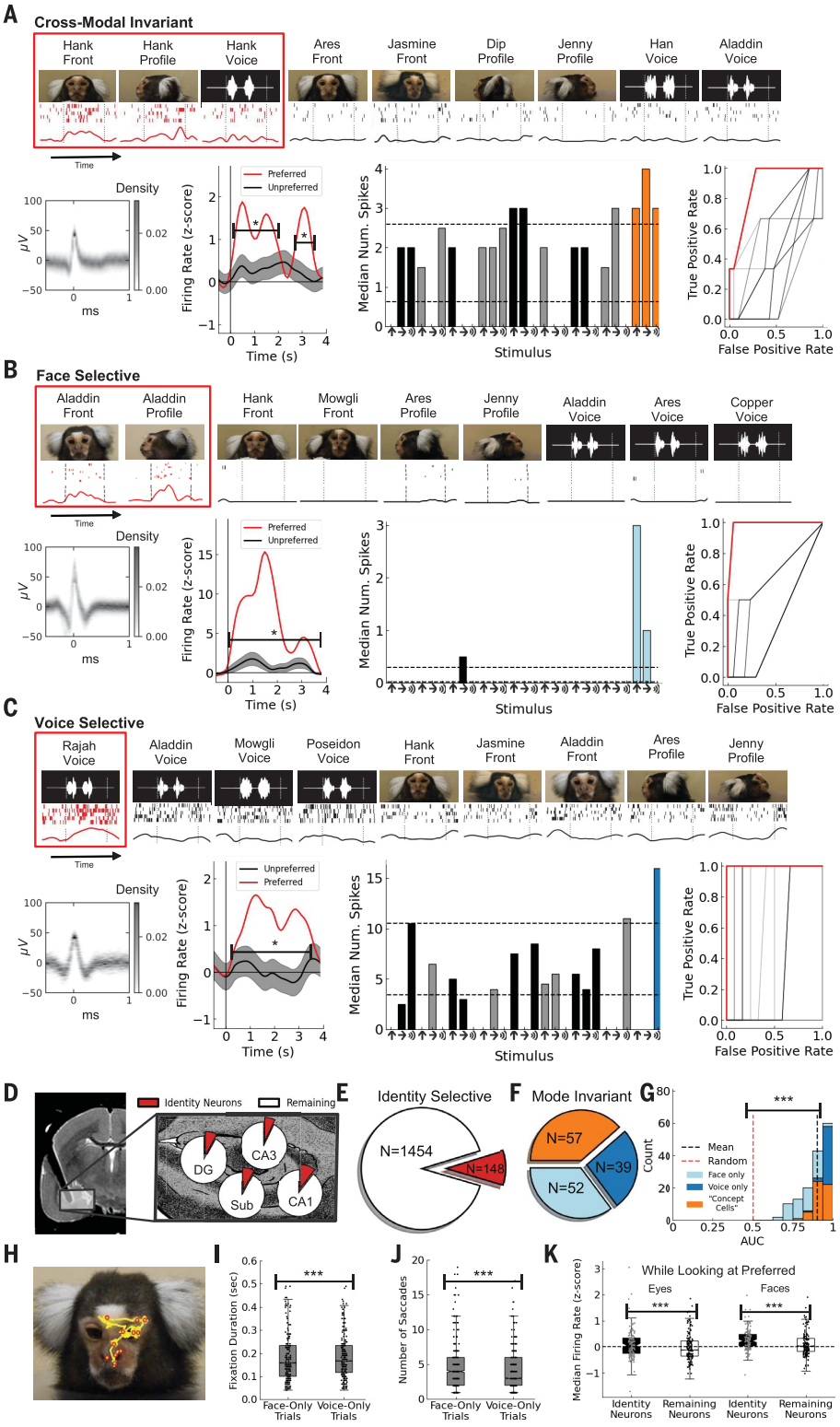

Fig. 1. Putative concept cells in marmoset hippocampus. (**A** to **C**) Top row: Subset of stimuli shown above raster and peristimulus time histogram (PSTH). Bottom row: Spike waveform density; normalized PSTH to all stimuli (preferred: red, nonpreferred: black), indicated are time points that show significant difference ($P < 0.05$); median number of spikes for unimodal stimuli (gray/black indicate nonpreferred individuals); ROC curve (shuffled controls shown in black). Exemplar identity neurons responding selectively to the face and voice of a preferred conspecific (red) (A), the face only (B), and the voice only (C) are shown. (**D**) Anatomical distribution of identity neurons (red) in hippocampal subfields relative to neurons remaining that responded to any stimulus (white). Black shadow indicates the electrode array track with magnetic resonance imaging distortion artifact. DG, dentate gyrus; Sub, subiculum. (**E**) Pie chart showing the abundance of identity neurons in red with the number of remaining neurons that qualified for the ROC selectivity analysis in white. (**F**) Mode distribution of identity neurons. Modes included face (light blue), voice (dark blue), and both (orange). (**G**) Histogram showing the distribution of AUCs comparable with red ROC curves in (A) to (C). Colors are as in (F). Black dotted line is the mean, and red dotted line is the mean of 10,000 random shuffles of the labels. (**H**) Exemplar eye movements (yellow) with fixations indicated (red). (**I**) Distribution of eye fixation durations for unimodal trials. (**J**) Distribution of apparent saccade number for unimodal trials. (**K**) Distribution of median firing rates while observer was looking at eyes (left) and face (right) for identity neurons (black) versus remaining neurons (white). Significant median differences, ***$P < 0.001$.

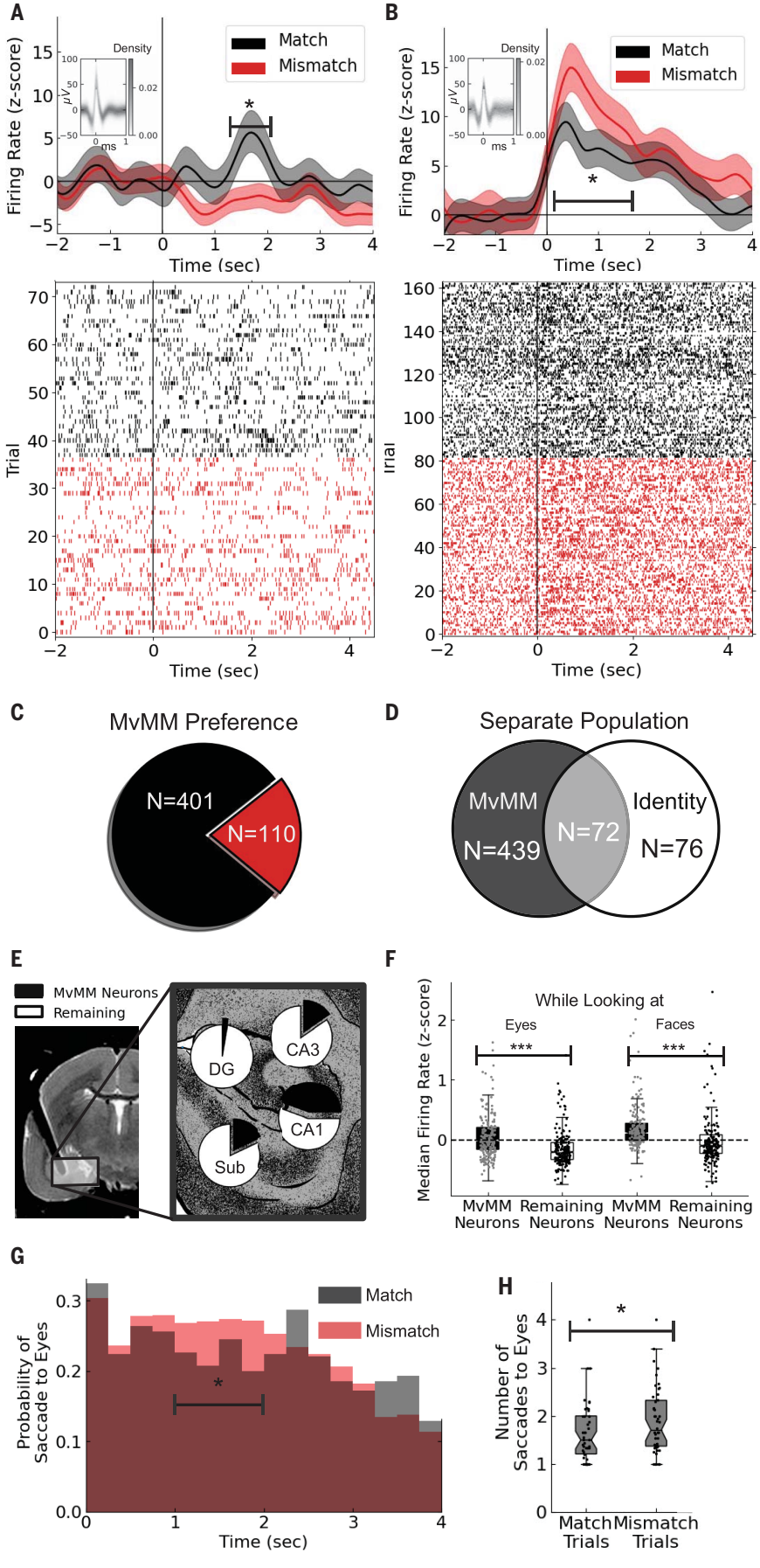

Fig. 2. Single neurons in hippocampus represent multiple individuals. (**A** and **B**) PSTH normalized by the prestimulus baseline (top) and spike raster (bottom) for two exemplar MvMM neurons. Black indicates match trials, and red indicates mismatch trials. Vertical line indicates stimulus onset. Inset shows spike waveform density. Significant time points, *$P < 0.05$. Exemplar neuron with higher firing rate for match (A) and mismatch (B) trials. (**C**) Pie chart showing the number of neurons that responded significantly more for match (black) or mismatch (red) trials. (**D**) Venn diagram showing the number of MvMM neurons (black) in common with identity neurons (red). (**E**) Relative abundance of MvMM neurons in each hippocampal subfield. (**F**) Distribution of median firing rate while looking at the eyes (left) and face (right) for MvMM neurons (black) versus remaining neurons (white). (**G**) Probability density of saccadic eye movements directed toward the eyes for match (black) and mismatch (red) trials. Indicated are the time points in (H). (**H**) Distribution of apparent number of saccades to eyes. Significant median differences, *$P < 0.05$.

neurons exhibited significant differences in median firing rate for a specific category ($P < 0.05$), which we labeled as predictive time bins (fig. S4). This algorithm was applied to all neurons in the population, not only those classified as identity-selective or MvMM neurons. We first identified predictive time bins selective for specific individuals when observing their face or voice. A pair of exemplar neurons that exhibited separate predictive time bins for two different individuals is shown (Fig. 3A and fig. S5). Out of 2358 hippocampal neurons, 1634 (69.3%) exhibited at least one identity-specific predictive time bin, with most exhibiting predictive time bins for two or more individuals (Fig. 3B). Identity-specific predictive time bins exhibited a mean AUC (0.802 ± 0.003) that was significantly above chance ($P < 0.001$, $N_{bins} = 3958$) (Fig. 3C). Neurons that had identity-specific predictive time bins exhibited a significantly greater median firing rate when subjects were looking at the face of a preferred individual ($P < 0.001$) (fig. S6A), although significant suppression was observed relative to

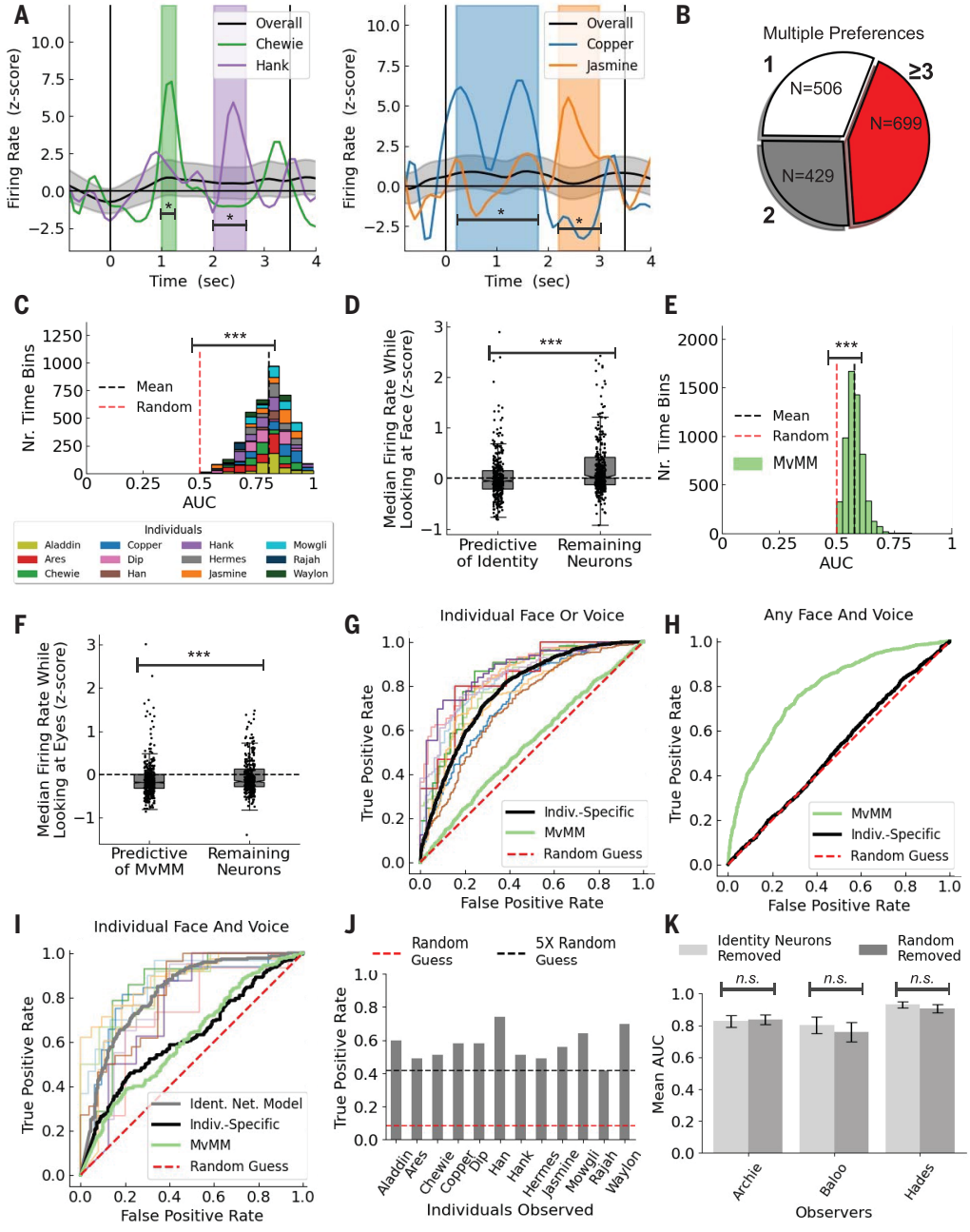

Fig. 3. Cross-modal decoding of identity. (**A**) PSTH of two exemplar predictive neurons. Colored traces average over trials involving preferred individual, and the gray shaded regions indicate 95% confidence intervals. Colored regions indicate identity-specific time bins. (**B**) Pie chart showing number of identity-specific predictive neurons that prefer one (white), two (gray), and three or more individuals (red). (**C**) Histogram showing AUC distribution of identity-specific time bins with colors indicating preferred individuals in legend. Dotted lines indicate the mean (black) and the control (red). (**D**) Distribution of median firing rates while the observer was looking at the face for the identity-specific predictive neurons compared with the remaining neurons. (**E**) Histogram showing AUC distribution of MvMM time bins. Dashed lines indicate the mean (black) and the control (red). (**F**) Distribution of median firing rates while the observer was looking at the face for the MvMM predictive neurons compared with the remaining neurons. (**G**) ROC curves for the detection of face or voice of individuals. Firing rates were considered from MvMM time bins (green, AUC = 0.536) and identity-specific time bins (black, AUC = 0.779) similarly averaged over individuals. Thinner colored lines indicate individuals as in (C). (**H**) ROC curves for the detection of match trials. Firing rates from MvMM time bins (green, AUC = 0.782) and from identity-specific time bins (black, AUC = 0.516). (**I**) ROC curves for the detection of both face and voice of individuals from same 19 recording sessions as in (G) and (H). Firing rates from MvMM time bins (green, AUC = 0.615), identity-specific time bins (black, AUC = 0.622), and the INM (gray, AUC = 0.818) are similarly averaged over individuals. Results of the INM for individuals are shown by thin lines colored as in the legend of (C). Red dotted line indicates random as in (G) and (H). (**J**) Bar plot showing true positive rates predicted by a winner-take-all model that considered predictions from the INM specific to 12 individuals. Indicated is the mean of the shuffled labels (red) and 5× that value (black). Bar plots summarize the trials from the testing sets of 33 recording sessions ($N_{trials} = 454$). (**K**) Bar plot showing mean AUC with identity neurons removed (light gray) versus the control randomly removing an equal number of bins from the remaining cells (dark gray). Uncertainty indicates 95% confidence of the mean. No significant difference was observed across recording sessions for any of the three qualifying subjects (Archie, $P = 0.81$, $N_{identities} = 14$; Baloo, $P = 0.58$, $N_{identities} = 9$; Hades, $P = 0.50$, $N_{identities} = 12$). ***$P < 0.001$; n.s., not significant.

the other neurons when normalizing by the background, which was averaged from t = −0.8 s to t = −0.3 s (P < 0.001) (Fig. 3D). We applied the same algorithm to test for predictive time bins that distinguished MvMM trials and found a similar result (Fig. 3E), with 1455 neurons exhibiting MvMM predictive time bins. Neurons with predictive time bins for MvMM exhibited a significantly greater median firing rate when subjects looked at the face (P < 0.001) (fig. S6B), although significant suppression was observed relative to the

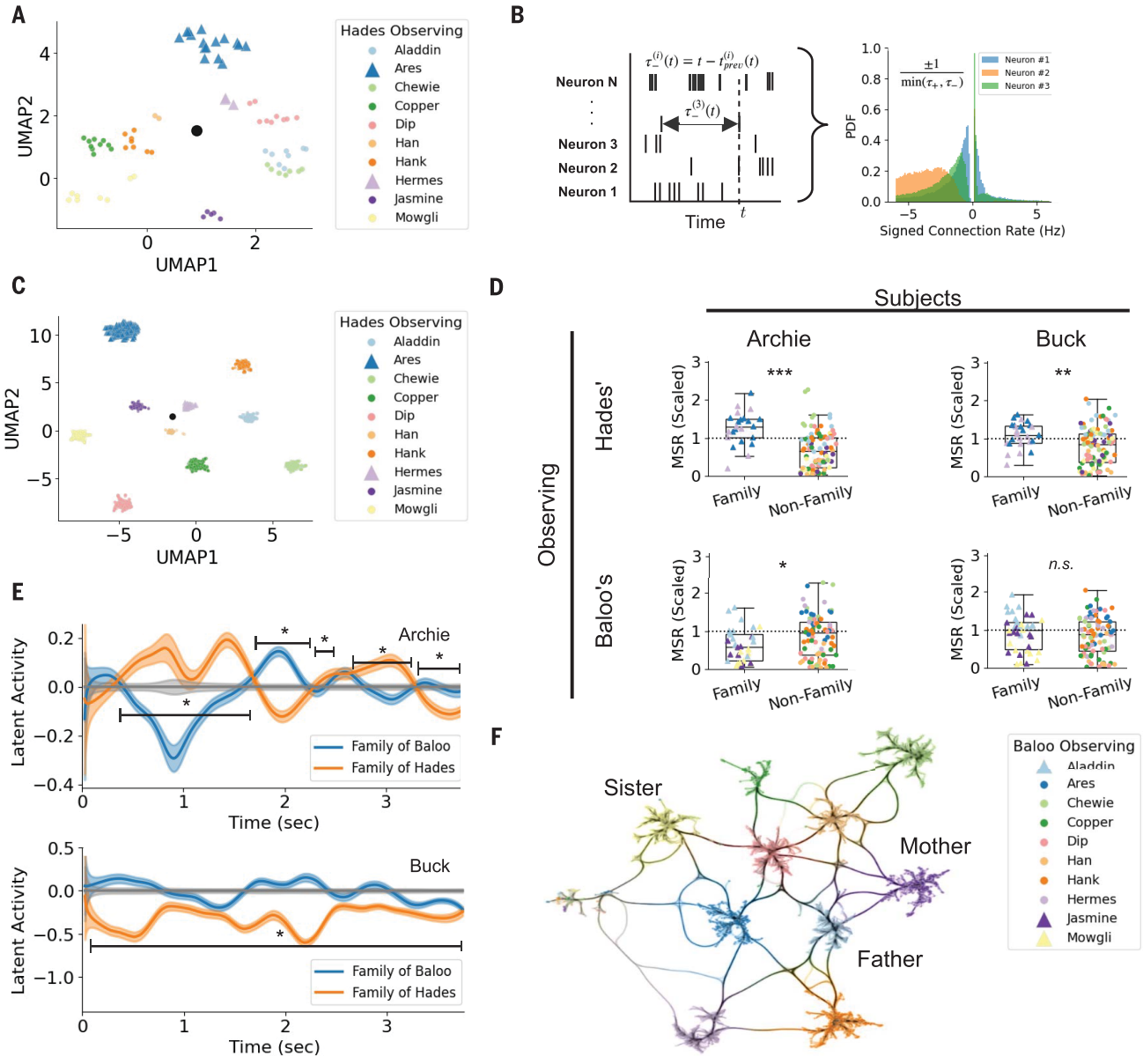

Fig. 4. Cross-modal representation of identity using rate and event codes. (**A**) Two-dimensional manifold projection of our rate-coded representation computed from firing rates of identity-specific time bins. One identity-match trial was equivalent to one presentation of the stimulus as face and voice matched. Each identity-match trial represented a different, randomly selected face and/or voice stimulus. Firing rates were computed from each identity-specific predictive time bins and then concatenated into a feature vector for each identity-match trial. This feature vector was then projected onto the manifold and plotted as one symbol per identity-match trial in one of the scatter plots. Indicated is the mean (black). Colors in legend correspond to individuals. UMAP, uniform manifold approximation and projection. (**B**) Schematic illustrating the hindsight delay to a given neuron (left), used to generate histograms of signed connection rates to three neurons (right). PDF, probability density function. (**C**) Two-dimensional manifold projection of our event-coded representation of identity computed as the manifold projection of signed connection rates of all neurons in the same exemplar recording session. One symbol represents one spike. Indicated is the mean (black). (**D**) Boxplots of MSR showing significantly different values when subjects observed family of other subjects. Shown is Archie observing family of Hades (top left, P < 0.001, $N_{identities}$ ≥ 23), Buck observing family of Hades (top right, P = 0.003, $N_{identities}$ ≥ 26), Archie observing family of Baloo (bottom left, P = 0.017, $N_{identities}$ ≥ 30), and Buck observing family of Baloo (bottom right, P = 0.828, $N_{identities}$ ≥ 37). Significance was computed according to Student's t test. (**E**) Latent activity averaged over all recording sessions from subjects Archie (left) and Buck (right). Colors indicate average over the family of Baloo (blue) and Hades (orange) relative to all conspecifics (gray). Shaded regions indicate 95% confidence of the mean estimated through bootstrap. (**F**) Graph of connections bundled between individuals. Triangles in legend indicate family members as in (A) and (C).

other neurons when normalizing by the same background ($P < 0.001$) (Fig. 3F). Instances of face and eye viewing were highly variable and not limited to the timing of predictive time bins, which suggests that attentional effects from visual behavior were not likely driving neural activity during these periods (fig. S7). We observed considerable overlap between neurons with identity-specific and MvMM predictive time bins because 82.2% ($N = 1196$; fig. S8A) exhibited predictive time bins in both analyses.

Identity network model

We developed a stable neural decoder by combining the firing rates of predictive time bins using an ensemble of gradient-boosted decisions trees (*21*). When using identity-specific time bins, we could reliably decode the identity of all marmosets when subjects observed their face or voice (accuracy: 77.4%) (Fig. 3G). The same approach could successfully decode MvMM trials when using MvMM time bins (accuracy: 75.7%) (Fig. 3H). The two kinds of decoders used mostly different time points, with only 24.6% ± 1.5% of identity-specific time bins overlapping with MvMM time bins within the same neurons (fig. S8B).

To test whether the same population could represent multiple cross-modal identities, we developed the identity network model (INM), which integrates these two decoding approaches. The first approach was identical to the identity-specific decoder described above, which resulted in accurate decoding for each individual's face or voice. The second approach classified MvMM trials as either match or mismatch but was anonymized to individual identity. Our INM combined these two approaches to achieve cross-modal decoding of individual identity (fig. S9). This combination was critical because the identity-specific predictive population was only accurate for individual identity but performed poorly for classifying MvMM (Fig. 3G), whereas the MvMM predictive population was the opposite (Fig. 3H). When combined across individuals, the INM successfully decoded the cross-modal identity of all 12 individuals (accuracy: 84.5%) (Fig. 3I). Decoding performance tested at least 5× above chance when distinguishing all individuals (Fig. 3J and fig. S10).

Because identity neurons were included in decoding, we investigated whether their explanatory contribution was disproportionate to their sparse distribution. We compared INM performance when these neurons were removed from the analysis and separately used only in the analysis versus an equal number of other neurons. We observed no significant effect on decoding performance despite the consideration of only individuals preferred by identity neurons (Fig. 3K and figs. S11 and S12), which suggests that these highly selective neurons are no more notable for decoding the cross-modal identity of familiar individuals in the hippocampus than other neurons in the same population. Furthermore, no significant effect of identity neurons on decoding was demonstrated at a larger 5-SD response threshold.

Social category representations in hippocampus

An individual's identity is also coupled to their social relationships, such as their family. To test whether hippocampus encodes categorical attributes of social identity, we applied nonlinear dimensionality reduction techniques (*22*). Using mean firing rates consistent with studies of face and voice processing in the primate brain (*23*), we first verified that these reduction techniques were capable of separating the stimulus categories at multiple probe locations along the anterior-posterior axis (fig. S13). We next replicated the findings of the INM using the same identity-specific predictive time bins for marmoset faces and voices drawn from the entire hippocampal population and showed that manifold projections similarly separated individuals (Fig. 4A and fig. S14), including for different subpopulations of neurons (fig. S14G). This suggests that cross-modal identity representations are evident in the population activity of marmoset hippocampus.

To investigate whether representation of identity can be described by the relative timing of spikes, we computed manifold projections of spike times recorded during identity-match trials (Fig. 4B, left) using parameterless signed connection rate features. The signed connection rate from one neuron to another describes how they interact, which reveals statistical distributions specific to any given pair of neurons (Fig. 4B, right)—a facet of neural activity distinct from the firing rate of any single neuron. Each spike was concatenated into a feature vector, which was then projected onto the manifold as is shown by one symbol (Fig. 4C). The feature vector was computed as the signed connection rate to each neuron at each observation time. Each observation time was a spike time of the neuron with the greatest number of spikes. Results using this event-coded measure revealed excellent separability for identity-match trials (Fig. 4C), which thereby replicated the effect observed with the INM using a distinct facet of neural activity.

We next investigated whether social categories other than identity may likewise be represented in event-coded hippocampal activity. We tested whether representations of other marmosets' family members were distinct from nonfamily members for the two marmosets whose families were not included in the stimulus sets using two distinct quantifications of manifold projections, although the pattern was consistent for all subjects. First, results revealed a significant difference in the mean square range (MSR) of the manifold projections along this category boundary (Fig. 4D and fig. S15A), which suggests that a larger event-coded state space was occupied while observing family members (fig. S15B). Although these projections were supervised, the clustering that emerged on the basis of respective social relatedness was unsupervised. Second, we computed the unsupervised latent firing rate as the manifold projection of the absolute value of signed connection rate. Although individual identities did not separate (fig. S16A), we found trajectories that appeared stable in time and comparable across trials (fig. S16B). The motion of mean latent firing rate significantly separated social categories at multiple time points for all subjects (Fig. 4E and fig. S16, C and D). Together, these results demonstrate that neural representations of social identity in primate hippocampus are not only invariant to the sensory modality and comparable over time (fig. S17) but that low-dimensional manifolds (Fig. 4F) can describe relationships between different social categories (e.g., individual identity, family groups).

Conclusions

We showed that the cross-modal identity of multiple conspecifics is represented in the primate hippocampus. Although we identified putative concept cells similarly to human studies (*9*, *12*), we discovered that this population of highly selective neurons is not the only mechanism for representing concepts of individuals. Rather, both single neurons and the broader population in hippocampus encode cross-modal identity of multiple conspecifics, similar to what has been reported for objects (*24*), which suggests that the sparse representations of concept cells may not be the only mechanism to represent semantic memory in hippocampus. An important caveat to these findings, however, is that our criteria for determining the responsiveness of putative concept cells differed somewhat from previous studies in humans (*9*–*11*), as described above. It is possible, therefore, that the concept cells in marmoset and human hippocampus are not strictly analogous. Ultimately, whether these neurons are equivalent is not determined solely by the physiological properties used to classify them in analyses but their functional role in memory. To directly address this issue, comparative experiments examining the computational contributions of concept cells in hippocampus across species during memory are critical, as such data are currently lacking. In addition to these findings at the single-neuron level, a population-level code representing not only the cross-modal identity of multiple familiar individuals but information pertinent to social categories was likewise reported in this study. Information from both the putative concept cells and those neurons

that encoded multiple identities were integrated, which suggests that cross-modal identity in hippocampus is evident in the ensemble activity of this keystone brain structure. Similar to the role of hippocampus in other contexts (fig. S18) (*25*), the cross-modal identity representations revealed in this experiment may support a learned schema that here applies to social identity (*26*, *27*). That these experiments were performed in a highly constrained context limits our ability to determine whether such a schema would indeed be leveraged in more naturalistic contexts during which social decisions on the basis of conspecifics' identity are made continuously (*6*, *8*). The presence of unimodal representations of identity in the primate frontal and temporal cortex (*2*, *8*, *28*), amygdala (*5*, *29*), and the medial temporal lobe (*30*) and representations of social dominance in the amygdala (*31*) may reflect an integrative social recognition circuit in which substrates in the broader network play distinct but complementary roles that collectively govern natural primate social brain functions (*32*).

REFERENCES AND NOTES

1. R. M. Seyfarth, D. L. Cheney, in *The Evolution of Primate Societies*, J. Mitani, J. Call, P. M. Kappeler, R. Palombit, J. B. Silk, Eds. (Univ. Chicago Press, 2012), pp. 629–642.
2. C. Perrodin, C. Kayser, N. K. Logothetis, C. I. Petkov, *Curr. Biol.* **21**, 1408–1415 (2011).
3. P. Belin, C. Bodin, V. Aglieri, *Hear. Res.* **366**, 65–74 (2018).
4. J. Sliwa, A. Planté, J.-R. Duhamel, S. Wirth, *Cereb. Cortex* **26**, 950–966 (2016).
5. U. Rutishauser *et al.*, *Curr. Biol.* **21**, 1654–1660 (2011).
6. M. C. Rose, B. Styr, T. A. Schmid, J. E. Elie, M. M. Yartsev, *Science* **374**, eaba9584 (2021).
7. L. Chang, D. Y. Tsao, *Cell* **169**, 1013–1028.e14 (2017).
8. R. Báez-Mendoza, E. P. Mastrobattista, A. J. Wang, Z. M. Williams, *Science* **374**, eabb4149 (2021).
9. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, *Nature* **435**, 1102–1107 (2005).
10. R. Quian Quiroga, A. Kraskov, C. Koch, I. Fried, *Curr. Biol.* **19**, 1308–1313 (2009).
11. I. V. Viskontas, R. Q. Quiroga, I. Fried, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21329–21334 (2009).
12. R. Q. Quiroga, *Nat. Rev. Neurosci.* **13**, 587–597 (2012).
13. R. Quian Quiroga, *Hippocampus* **33**, 616–634 (2023).
14. H. S. Courellis *et al.*, *PLOS Biol.* **17**, e3000546 (2019).
15. F. Mormann *et al.*, *J. Neurosci.* **28**, 8865–8872 (2008).
16. J. Minxha *et al.*, *Cell Rep.* **18**, 878–891 (2017).
17. M. Rigotti *et al.*, *Nature* **497**, 585–590 (2013).
18. C. T. Miller *et al.*, *Curr. Biol.* **32**, R482–R493 (2022).
19. D. Kumaran, E. A. Maguire, *J. Neurosci.* **27**, 8517–8524 (2007).
20. M. Fyhn, S. Molden, S. Hollup, M.-B. Moser, E. Moser, *Neuron* **35**, 555–566 (2002).
21. T. Chen, C. Guestrin, in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, 2016), pp. 785–794.
22. E. H. Nieh *et al.*, *Nature* **595**, 80–84 (2021).
23. W. A. Freiwald, D. Y. Tsao, *Science* **330**, 845–851 (2010).
24. T. P. Reber *et al.*, *PLOS Biol.* **17**, e3000290 (2019).
25. P. Baraduc, J.-R. Duhamel, S. Wirth, *Science* **363**, 635–639 (2019).
26. J. Sliwa, J.-R. Duhamel, O. Pascalis, S. Wirth, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1735–1740 (2011).
27. I. Adachi, R. R. Hampton, *PLOS ONE* **6**, e23345 (2011).
28. D. Y. Tsao, M. S. Livingstone, *Annu. Rev. Neurosci.* **31**, 411–437 (2008).
29. K. M. Gothard, F. P. Battaglia, C. A. Erickson, K. M. Spitler, D. G. Amaral, *J. Neurophysiol.* **97**, 1671–1683 (2007).
30. S. M. Landi, P. Viswanathan, S. Serene, W. A. Freiwald, *Science* **373**, 581–585 (2021).
31. J. Munuera, M. Rigotti, C. D. Salzman, *Nat. Neurosci.* **21**, 415–423 (2018).
32. W. A. Freiwald, *Curr. Opin. Neurobiol.* **65**, 49–58 (2020).
33. T. Tyree, M. Metke, C. Miller, Cross-modal representation of identity in primate hippocampus, Dataset, *Dryad* (2022).

ACKNOWLEDGMENTS

We thank H. Courellis for assistance with data collection and D. Leopold for comments on a previous version of this manuscript. **Funding:** This study was supported by National Institutes of Health grant R01 NS109294 (to C.T.M.) and National Institutes of Health grant R01 DC012087 (to C.T.M.). **Author contributions:** Conceptualization: C.T.M.; Methodology: T.J.T., M.M., and C.T.M.; Investigation: M.M. and T.J.T.; Analysis: T.J.T.; Visualization: T.J.T.; Funding acquisition: C.T.M.; Supervision: C.T.M.; Writing – original draft: T.J.T. and C.T.M.; Writing – review & editing: T.J.T. and C.T.M. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials or are deposited at Dryad (*33*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf0460
Materials and Methods
Figs. S1 to S20
Table S1
References (*34*–*38*)
MDAR Reproducibility Checklist

Submitted 25 September 2022; resubmitted 30 May 2023
Accepted 1 September 2023
10.1126/science.adf0460

Cross-modal representation of identity in the primate hippocampus

Timothy J. Tyree, Michael Metke, and Cory T. Miller

Science **382** (6669), . DOI: 10.1126/science.adf0460

Editor's summary

There are numerous animal studies showing that single neurons encode the identity of conspecifics from social signals in different sensory modalities such as olfaction, vision, or audition. However, there has been no demonstration in a nonhuman animal that these unimodal signals are integrated into a cohesive cross-modal representation of individual identity. Tyree *et al*. performed single-neuronal recordings in marmosets presented with pictures and sounds from conspecifics (see the Perspective by Wirth). A subpopulation of neurons responded selectively to both the face and voice of individual animals, much like concept neurons in the human medial temporal lobe. Furthermore, animal identity could be successfully decoded from the neuronal population activity, along with aspects of social relationships such as being a family member. —Peter Stern

Science
AAAS

Supplementary Materials for

Cross-modal representation of identity in the primate hippocampus

Timothy J. Tyree, Michael Metke, Cory T. Miller

Corresponding author: Cory T. Miller, corymiller@ucsd.edu

Science **382**, 417 (2023)
DOI: 10.1126/science.adf0460

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Materials and Methods

Subjects

Four adult marmosets (2 male, 2 female) served as subjects in these experiments. All animals are socially housed with 2-8 conspecifics in the Cortical Systems and Behavior Laboratory at the University of California San Diego (UCSD). All animals housed in a cage are family members, as each cage comprises a pair-bonded adult male and female and 1-3 generations of offspring. The UCSD marmoset colony in the Miller Lab houses ~70 animals in 15 family groups in a single room with visual and acoustic access between cages. All procedures were approved by the Institutional Animal Care and Use Committee at the University of California San Diego (S09147) and follow National Institutes of Health guidelines. A total of 47 recording sessions were performed with these subjects over the course of the experiment and analyzed here.

The total number of single units recorded from marmoset hippocampus totaled N=714 in Archie, N=822 in Baloo, N=212 in Buck, and N=610 in Hades (Figure S3B). All four subjects were considered equally in the identity neuron analysis and the MvMM neuron analysis (Figures 1, 2). All subjects were considered in the predictive time bin analysis (Figure 3) except for Buck due to his low count of single units across his 13 recording sessions. For the manifold projection analysis (Figure 4), all subjects were considered while they observed families that had at least two family members from amongst the cohort of individuals shown. All subjects were wild type common marmosets (Archie: male deceased at 3 years 6 months, Baloo: female deceased at 3 years 2 months, Buck: male deceased at 6 years and 4 months, Hades: female deceased at 1 year and 10 months).

Experiment Design

Neurophysiological recordings were performed while subjects were head and body restrained in our standard marmoset chair (*34*) Visual stimuli were presented on an LED screen from a BenQ monitor 1080 positioned 24 cm in front of the animal. Acoustic stimuli were presented at 70-80 dB SPL from a speaker positioned below the monitor (Figure S19). All behavior was collected in an anechoic chamber illuminated only by the screen, which had a dynamic range from 0.5 to 230 cd/m^2, with luminance linearity verified by photometer. Stimulus presentation was controlled using custom software and eye position was monitored by infrared camera tracking of the pupil. For hardware, calibration, and validation see previous work in the lab (*34*). During the recording session, we were blind to the randomized presentation of four hundred stimuli per recording session, which was the maximum duration that was practical for a subject.

Subjects initiated trials by holding fixation of gaze for 100ms at a center fixation dot on the screen, at which point stimulus presentation was initiated. The 150ms period immediately post-stimulus was discarded to account for the time for visual signals to propagate from the retina to the hippocampus. This latency has been measured to be in the range 100-200ms (*35*). This biophysical argument supports our estimate of the stimulus onset t=0 occurring 150ms after stimulus was presented. Unless otherwise specified, baseline firing rates were estimated from 500ms preceding t=0 excluding 300ms for anticipatory firing, as mentioned in the main text and later. Stimulus responses were initially measured by comparing the peristimulus baseline firing rate to firing rates averaged from the max of a 500ms sliding window from t=300ms to 3.5s, as mentioned in the main text and later.

Stimuli were divided amongst unimodal– face-only and voice-only– and cross-modal– identity match and identity mismatch– on a trial-by-trial basis. Up to twelve conspecifics were represented per stimulus set (min 10, max 12). Face stimuli comprised multiple examples of each individual marmoset from different head orientation.

All face and voice stimuli were pictures or audio recordings from animals housed in the same colony room as the subjects. Because the colony is housed in a single room in which all animals have visual and acoustic interactions with each other, we assumed that all animals have sufficient experience observing each other to be familiar with their respective individual identities. Each individual marmoset was represented in multiple distinct stimuli ($N_{stimuli}$=36.0±15.3) for each individual in each recording session across each of the three stimulus classes: face forward, face profile and vocalization. No single stimulus

was presented to subjects more than two times in a single test session. Monkeys with fewer than 10 presentations per individual in a recording session were not considered in any analysis. The stimulus duration of trials involving vocalizations (i.e. voice-only and cross-modal) necessarily varied because each "phee" call differed in duration (mean: 3.02±0.74s). The median face stimulus duration was 3.50 seconds (IQR: 2.78-3.51 seconds). The minimum face stimulus duration was 2.05 seconds and the maximum face stimulus duration was 4.46 seconds. Stimuli were presented in 10-trial blocks, with an inter-block active forage trial with juice reward to maintain attention. Each recording set was composed of 400 face and/or voice stimuli, split into 2 subsets.

All stimuli were composed of faces and/or voices of conspecific monkeys in our colony familiar to each subject. A total of 16 individual monkeys were represented overall (9 male, 7 female). Test subjects were not included in their own stimulus sets. Because our goal was to test for representations of individual identity rather than cross-modal perceptual integration of face/voice biomechanical movements (i.e. McGurk Effect) we presented subjects with static face stimuli so as not to introduce confounds that may emerge due to temporal misalignments of the face and vocalizations during the identity mismatch trials.

All face stimuli were photographs of monkeys from our colony taken while animals were in our standard marmoset chair with a light background behind them. The animals are trained to sit comfortably while a neck guard restricted their mobility. While seated, subjects could freely change head direction. Photographs of each subject were visually inspected and selected based on image quality and suitable representation of multiple head orientations (Figure 1A-C, S1). Photos used as stimuli were cropped to only show the neck guard and the face/head, so as to eliminate views of the rest of the body and chair.

All voice stimuli were marmoset "phee" calls comprising two pulses, the species-typical long-distance contact calls. Previous work has shown that marmosets are able to recognize the caller's identity when hearing "phee" calls (*36*). Recordings were made at 44.1kHz sampling rate while a monkey engaged in natural vocal interactions with a visually occluded conspecific in a soundproof chamber and hand-selected using custom code. Only examples with high SNR and minimal background noise were selected for stimuli.

All analyses were performed in Python unless otherwise indicated.

Surgical and neural recording details

The surgical procedure employed here has been described previously (*14*). Briefly, we performed an initial surgery to affix a post to the skull on each animal to restrain subjects' head during experiment preparation. Following recovery, a second procedure was performed to embed the drive housing and the electrode array for stable chronic electrophysiological recording. We implanted a 64-channel microwire brush array (MBA, Microprobes) either unilaterally or bilaterally into the hippocampus using preoperative MRI stereotaxic coordinates. Electrode locations were confirmed by postoperative MRI and histology. All surgeries were performed under sterile and anesthetized conditions. The implants were inserted 7-13 degrees of angle off the vertical using the medial sulcus as reference before the operation has taken place. Neural recordings were performed with an Intan 512ch Recording Controller system via an RHD2164 64-channel amplifier chip, sampled at 30kHz. Neurophysiology data was analyzed using Spyking Circus yielding across all recording sessions 2,358 isolated units, referred to as neurons in the main text and in the remainder of Methods and Materials. Standard procedures were employed to remove obvious recording errors, which resulted in less than 1% of trials being removed from the analysis *a priori*.
Statistical tests comparing median firing rates were Wilcoxon-Mann-Whitney tests because they make no assumption that specifies the distributions of its arguments.

Identifying Identity Neurons

Hippocampal neurons were tested for an invariant response to individuals in the face-only and voice-only trials using an ROC analysis similar to that described in human hippocampus (*9*). For each isolated single neuron we performed the analysis for all identities where at least 4 unimodal stimuli (either

face or voice but not both) were presented for each of the following three unimodal stimulus categories: face forward, face profile and voice.

The response of a neuron to a trial was the taken to be the maximum spike count in a 500 millisecond continuous sliding time window from t=0.3 seconds to t=3.5 seconds following stimulus onset at time t=0, as described above and in the main text. As in (*9*), the response of a neuron to a stimulus was the median response averaged over all presentations of the stimulus.

A neuron was considered responsive to a stimulus if its response to the stimulus was above the responsiveness threshold, which was determined as the sum of the mean baseline plus two standard deviations (s.d.) of the baseline, where the baseline was the number of spikes averaged over the times t=-0.8 seconds to t=-0.3 seconds, as described above and in the main text. This differs from the original study in humans (*9*), which used five s.d. instead of two, which was not practical in this study due to marmoset hippocampal neurons typically exhibiting larger baseline firing rates (Figure S2), for which five s.d. would have resulted in responsiveness thresholds that would only be evident in N=166 out of the 2,358 single units involved in this study (7.0%). At a five s.d. response threshold, zero neurons were face and voice invariant (0%), four neurons were selective for faces (2.4%), and twenty-six neurons were selective for voices (15.7%). We decreased the significance threshold from five s.d. (mean threshold: 32.30 Hz) to two s.d. (mean threshold: 16.80 Hz) when identifying putative 'concept cells' to support the same significance level for selectivity ($p<0.01$). For comparison with other studies, the mean stimulus response in marmosets was 11.98 Hz and the mean background firing rate was 6.47 Hz.

A neuron was considered cross-modal invariant to an individual if it was responsive to all three unimodal stimulus categories for that individual. If a neuron instead responded only to the voice of an individual, then it was considered voice-invariant. If a neuron instead responded to an individual for both the front facing and profile facing stimulus categories, then it was considered face-invariant.

As in (*9*), stimuli were considered in ROC selectivity analyses only if at least one neuron responded to it. Also as in (*9*), an above-threshold response to a stimulus of the preferred subject was considered a positive test. Significance of an ROC for a given subject was determined by comparison to 99 surrogate ROC curves, which resulted from randomly and independently shuffling the labels. An area under the curve (AUC) that surpassed that of all surrogates was considered significant ($p<0.01$). Neurons that met or exceeded these thresholds were necessary to determine selectivity for individual identity in marmosets.

If a neuron was determined to be invariant to an individual within a given mode or modes, then selectivity was determined using the same mode or modes for that same individual. That is, cross-modal invariant neurons were tested for selectivity using all three unimodal stimulus categories, face-only invariant neurons were tested for selectivity using only front facing and profile facing unimodal stimuli, and voice-only invariant neurons were tested for selectivity using only the voice.

Cross-modal invariant neurons that passed the ROC selectivity test of (*9*) were considered selective for the identity and were thus labeled as putative 'concept cells'. Because all voice-only unimodal stimuli were combined into a single stimulus category, voice-invariance would imply voice-selectivity for one identity if not for an additional statistical test that compared the median trial response to the voice stimuli of the preferred individual to that of all other individuals according to a one-sided Wilcoxon-Mann-Whitney test ($p<0.01$) with an above-threshold response constituting a positive prediction of the preferred individual. The comparable test was used to determine selectivity for the face-invariant neurons. The invariant neurons demonstrating selectivity were considered identity neurons.

Identifying MvMM Neurons

Determination of MvMM neurons was achieved by comparing the median response of a neuron to identity match trials to the median response of that same neuron to identity mismatch trials. If a neuron was responsive to either match or mismatch trials, then a statistically significant difference computed according to a Wilcoxon-Mann-Whitney test qualified a neuron as a MvMM neuron ($p<0.05$). Preference of a MvMM neuron to match or mismatch trials was subsequently determined by a one-tailed Wilcoxon-Mann-Whitney

test ($p<0.05$). Importantly, we did not preselect for neurons that were broadly stimulus driven, but focused analysis only during the median stimulus and compared activity between match and mismatch trials. This is reflected in the exemplar neurons selected for Figure 2. The match preferent neuron (Figure 2A) shows a difference in firing rate during presentation of the stimuli but is not broadly stimulus driven. By contrast, the mismatch preferent neuron (Figure 2B) exhibits stimulus driven activity as well as differential firing rate between the stimulus types.

To quantify the responsivity of MvMM neurons in a way specific to multimodal integration, we computed the multimodal index (Figure S3A). The multimodal index is a measure of responsivity of a neuron to the combination of two modes relative to the greatest of either mode presented independently. The multimodal index (MMI) is the quantity given by

$$MMI = \frac{r_{MvMM} - \max(r_{face}, r_{voice})}{r_{MvMM} + \max(r_{face}, r_{voice})},$$

where r_{mode} is the mean response averaged over all trials of a neuron responding to a given mode.

Identifying predictive time bins

Hippocampal neurons were analyzed in terms of their firing rate response during time bins that we identified as candidate time bins. For each neuron, our procedure consisted of three stages. The first stage was to generate a large list of time bins of varying duration using an extension of a sliding window approach. The second stage identified a subset of time bins as having a general ability to distinguish trials. We required this subset to be mutually disjoint. Candidate time bins resulted from the third stage, which varied each time bin independently according to our refining procedure.

The first stage extended the sliding window approach by using 200ms time bins evenly distributed between 0 and 3.6 sec, the maximum stimulus duration (Figure S4A). Time bins of duration greater than 200ms were constructed by joining adjacent time bins, leading to a maximum allowed time bin duration of 3.6 seconds. A general ability to weakly distinguish trials was determined by splitting the training trials according to three-fold stratified cross-validation and then computing the training AUC of each fold (Figure S4B). Training AUC was initially computed from the ROC curve that resulted from an above-threshold firing rate response determining a positive trial. Separately, training AUC was computed from a below-threshold firing rate response as determining a positive trial. In either case, if the training AUC was greater than chance (AUC>0.5) for all three folds, then the time bin was retained for stage two. The same convention for *above* versus *below* firing rate response as determining a positive trial was used for stage two and for stage three. All population-level decoders were blind to this convention of sign.

The second stage selected a disjoint set of candidate time bins, optimizing for their ability to distinguish trials by maximizing the mean AUC averaged over the same three folds. To achieve this, time bins were selected in decreasing order of their mean AUC and included only if doing so maintained the disjointness of time bins.

To reduce the effect of discretizing the trial into time bins, the third stage refined the resulting disjoint set by considering a number of random perturbations of each remaining candidate time bin and keeping only the optimal perturbation. The random perturbations shifted the start times and the end times independently by a random amount identically sampled from the normal distribution with zero mean and standard deviation equal to the duration of the unperturbed time bin. We generated a sample of N=100 perturbed time bins and removed those with a duration <10ms. Perturbations were additionally removed if they exhibited a start time before stimulus onset t=0 or if they exhibited an end time after t=3.6 seconds. A worsening AUC in any of the folds resulted in rejection of the given candidate time bin.

If any of the resulting training AUC values were smaller than that of the unperturbed time bin, there that perturbation was removed from consideration. The overall training AUC was computed for each perturbation using all training trials together. The perturbed time bin with the largest overall training AUC

was kept instead of the unperturbed time bin. Perturbed time bins were allowed to overlap with other remaining time bins, thereby relaxing the condition of disjointness for the sake of parallelizability, which is statistically valid because zero spike times in the training set appear in the testing set and the decoder makes no assumption of independence of features. A flowchart summarizes the time bin refinement procedure (Figure S4C).

If no perturbations remained under consideration, then the unperturbed time bin was kept from stage two. Any remaining candidate time bins were considered predictive only if they presented a statistically significant difference in median firing rate for the true (e.g. identity match) training trials compared to the false (e.g. identity mismatch) training trials. Significance was determined according to $p<0.05$, where p was the statistic computed as the mean p-value resulting from a Wilcoxon–Mann–Whitney test conducted over the training trials averaged over five stratified cross-validation folds over training, which was a sufficient statistic in the sense that all time bins with $p<0.05$ also exhibited a statistically significant difference in median value at the same level of significance according to a Wilcoxon-Mann-Whitney test conducted over all MvMM trials. This procedure provided the features used in our population-level decoders. Data and code are made available to the reader (see Author Contributions).

Training the population-level neural decoders

Population-level decoders were trained on the training trials before computing predictions for the separate testing trials. Decoders were trained and tested on a Quadro RTX 5000 GPU typically in less than five seconds of runtime.

The population-level decoders trained using firing rates directly as inputs. Neither translating nor scaling of the firing rates was performed, as the decoders were both location and scale invariant (*21*). The prediction was estimated by the weighted average of values returned by an ensemble of gradient-boosted decision trees relative to a default value of one half (controlled by base_score in Table S1). For each training epoch, at least 25 decision trees were trained (controlled by num_parallel_tree). While a unique solution exists for a given decision tree, a heuristic algorithm was used to approximate the unique solution using the quantile method of (*37*).

Decision trees were trained to minimize the binary cross-entropy loss function (equivalently, to maximize likelihood) at the ensemble-level by considering only a fraction of the training trials (controlled by subsample). Decision node rules considered only a fraction of the input firing rates (controlled by colsample_bynode) to determine placement of its weight. The weight of a node was limited to a certain amount (controlled by max_delta_step). The complexity of the decision node rules was further limited using linear and quadratic regularization (controlled by reg_alpha and reg_lambda in Table S1, respectively).

Each decision tree was gradient boosted in the sense that nodes were recursively added in accordance with an estimate of the gradient of a training loss computed at the ensemble-level. If inserting a decision node failed to improve the loss by a sufficiently large amount (controlled by gamma), then that decision node was removed from the tree. To further limit structural complexity, the maximum tree depth was set to no more than five decisions (controlled by max_depth). The weight for a new decision tree was scaled down by a factor (controlled by learning_rate). Training terminated for a given decision tree when the total weight for the next decision node was smaller than a certain amount (controlled by min_child_weight). After all decision trees terminated training, the training epoch was complete. After a fixed, predetermined number of training epochs, the ensemble terminated training. Then, predictions were computed for the testing trials (Figure S9A). Predictions were used to evaluate the predictive ability of a given set of one or more predictive time bins in terms of AUC.

Determining hyperparameter settings for the population-level neural decoders

The parameter settings for our population-level neural decoders resulted from a series of coarse grid searches each conducted over a wide range of settings for one pair of hyperparameters at a time. Each parameter setting considered five-fold stratified cross-validation involving the training trials only with the

goal of maximizing mean testing AUC. Early stopping was used during this tuning procedure, which supported a minimum 60 training epochs for the match vs mismatch (MvMM) predictive population and a minimum 67 training epochs for the identity-specific predictive population as sufficient according to early stopping. By increasing the number of training epochs, stability of performance became immediately apparent for up to 500 epochs for both MvMM and identity-specific decoders. We made no use of early stopping anywhere else apart from the hyperparameter tuning procedure described here.

This hyperparameter tuning procedure was conducted only on the training trials for Archie observing Waylon in one recording session from subject, Archie (session #8). Archie (male) and Waylon (female) were not family members– though they likely knew each other in the colony. These training trials (from session #8) were complementary to testing trials from no more than one of the multiple recording sessions summarized in Figure 3. The hyperparameter settings that resulted are reported in Table S1.

Summarizing testing performance from multiple predictors

Population-level decoders were trained as MvMM or identity-specific predictors for each individual identity in each recording session involved in Figure 3. To account for variations in prediction magnitude between decoders, predictions were scaled linearly to a maximum value of unity before combining ROC traces in the multiple recording sessions summarized in Figures 3G-I,K and S11-12. No such scaling was involved with the multiclass predictions reported in Figures 3J and S10.

Sampling trials for multiple predictive populations from the same recording session

For a given recording session, the following criteria were respected while partitioning testing trials from training trials involving the identity network model (INM) discussed in the main text. Testing trials for the INM were also testing trials for both the MvMM decoder and the identity-specific decoders. Because stimuli involving individuals were sampled uniformly, the frequency of a given individual could be small for a given recording session. To account for this, individuals were considered only if they exhibited at least forty appearances in a given recording session.

Because of the uniform nature of our uniform random sampling of trials over the larger space of cross-modal stimuli, each recording session had relatively few trials involving both the face and the voice of a particular individual. This resulted in far more negative trials being presented to the observer relative to the number of true trials for the INM. This was also the case for both the MvMM decoders and the identity-specific decoders reported in Figures 3 and S10-12. All three binary classification tasks had balanced samples randomly selected, which were then randomly shuffled before 30% were randomly selected to be testing trials. The remaining 70% of trials were considered for training. Unbalanced sampling in the training set was accounted for by scaling the positive weights by a factor of 5 for the MvMM decoders and 100 for the identity-specific decoders. Decoders involved in Figure 3 used 200 training epochs, all of which were used in testing decoder performance except the first training epoch. The only exception was the identity-specific decoders involved in evaluating the INM for the winner-take-all model in Figures 3J and S10, which considered all 500 training epochs.

Decoding multiple identities using a winner-take-all model

We used the winner-take-all model to predict the identities of multiple individuals shown during identity match and face-only trials. The twelve individuals summarized (Figure 3J) have their detailed testing performance reported (Figure S10). The winner-take-all model predicted the correct identity with an overall testing accuracy of 91.0% (N_{trials}=454). For a given recording session, the following procedure was performed to generate the predictions for the winner-take-all model. First, we identified all identities involved in a sufficient number of identity match trials ($N_{trials}\geq12$). All identity match trials involving the identities identified were shuffled and 30% were randomly selected as testing trials to be withheld from training with the remaining 70% of trials.

We considered predictions of our INM to approximate a predicted probability that a given trial from the testing set involved the given identity. The presence of the individual was modeled using the decoder outputs in the winner-take-all model if the INM had the sufficient number of predictive time bins available. After repeating this procedure for all individuals in the recording session, the predicted identity of the winner-take-all model corresponded to that of the maximum predicted value (Figure S9B).

Quantifying relative contribution of identity neurons in decoders of preferred identities

To investigate the possibility of identity neurons exhibiting any clearly observable significance in the INM at the population-level, we removed all identity neurons from consideration and recomputed the testing predictions of Figure 3I for each individual that was statistically preferred by an identity neuron. After recording the testing AUC, we repeated a comparable procedure as a control that randomly removed an equivalent number of predictive time bins from any neuron that was not found to be an identity neuron. This control procedure was repeated many times ($N_{samples}$=200) and then averaged to estimate the mean control testing AUC, which was not significantly different from a normal distribution according to D'Agostino-Pearson's omnibus test ($p>0.05$, $N_{samples}$=200). The aforementioned control and test procedures were conducted using independent randomized samples.

ROC curves were computed with above-threshold values indicating a positive trial for the three observers with at least two family members amongst the identities presented. The INM appeared successful despite the removal of identity neurons independently for multiple observers (Figure S12). Removing identity neurons from the INM for all recording sessions involving one observer resulted in a mean testing AUC that was not significantly smaller than that of the control according to a one-tailed paired student's t-test that supposed identity neurons contributed more to decoding than other neurons. We independently replicated this same statistical insignificance of identity neurons at the population-level for multiple observer subjects ($p>0.05$, $N_{observers}$=3). This insignificance was consistent with a comparable analysis that made no assumption of normality, which suggested the median testing AUC was also not significantly smaller when all identity neurons were removed relative to the control ($p>0.05$, $N_{observers}$=3). It is uncertain whether this insignificance can be attributed to these identity neurons being observed in nonhuman primates, as no comparable predictive time bin analysis has ever been performed in humans to the knowledge of the authors.

Computing signed connection rate

Our event-coded representation relied on our signed connection rate measure, which we computed using our two primitive event measures. The first we referred to as the hindsight delay, $\tau_- > 0$, which is the amount of time since a given neuron has spiked. The second we refer to as the foresight delay, $\tau_+ > 0$, which is the amount of time until a given neuron will spike. A schematic illustrating the computation of the hindsight delay is shown (Figure 4B, left). A similar computation is found for the foresight delay by time inversion. If the given neuron has not yet spiked, then we take the hindsight delay to approach infinity. Similarly, if the given neuron was not observed to spike again, then we take the foresight delay to approach infinity. Note that our primitive event measures do not evaluate to non-positive real numbers.

The magnitude of our signed connection rate is the multiplicative inverse of the minimum of the hindsight delay and the foresight delay. Finally, we set the sign of our signed connection rate to be negative if the hindsight delay was used. Using the standard conventions of real analysis, our signed connection rate is now well-defined at all times for all neurons that exhibited at least two spikes. Equivalently, our signed connection rate was computed according to a real function of two variables

$$c(\tau_+, \tau_-) = \frac{\theta(\tau_- - \tau_+)}{\tau_+} - \frac{\theta(\tau_+ - \tau_-)}{\tau_-},$$

where $\Theta(x)=1$ if x is nonnegative, otherwise, $\Theta(x)=0$. We found a statistically significant correlation between the signed connection rate and firing rate during predictive time bins ($p<0.001$) according to Pearson's test for correlation (correlation: -0.223), Spearman's test for correlation (correlation: -0.639), and Kendall's test for correlation (correlation: -0.461). This is explained by signed connection rate being normalized for spiking rate by design. One can view the signed connection rate magnitude as a sample of one divided by the distribution of inter-spike intervals. As firing rate is defined by a local average over the same distribution of inter-spike intervals, a statistical correlation must be apparent, as was observed.

We evaluated our signed connection rate for every neuron at the spike times of the neuron that spikes the most over the recording session (i.e. the reference neuron). This was our attempt to measure how a single neuron "connects" with any other neuron. In doing this, we observed statistical distributions that appeared specific to a given neuron pair (Figure 4B, right), as discussed in the main text. We considered a given neuron to have an approximately symmetric signed connection rate if it exhibited no more than twice as many negative values as positive values in these statistical distributions.

Estimating manifold projections

We used uniform manifold approximation and projection (UMAP) to compute our manifold projections in Figure 4 of the main text, which presents descriptive manifold projections computed from predictive firing rate features and separately from our signed connection rate measure of spiking events. The same parameter settings on the same optimization algorithm was used for both rate and event-coded manifold projections. We used the identity-specific predictive time bins in the rate-coded representation. The rate-coded manifold projections considered neuron spikes from t=0 to 2 seconds after the stimulus onset. Similarly, the event-coded manifold projections considered neuron spikes from t=0 to 2 seconds after the stimulus onset. The average predictive time bin from the MvMM predictive population reported in Figure 2 was centered from t=0 to 2 seconds after the stimulus onset, with approximately half of predictive time bins ending earlier, which supports 2 seconds as a reasonable choice for the max time considered by the rate and event-coded manifold projections.

The UMAP algorithm was composed of two steps that can fruitfully be described as graph construction and graph projection (*38*). The graph was constructed from a given set of comparable observations. The graph was projected to a low-dimensional space of real numbers. The output was embedded in two to three dimensions for visualizations and statistical analyses. In the optimization procedure, five negative samples were selected for each positive sample. The minimum distance between two observations was set to 0.1Hz. The number of nearest neighbors was initialized to 50 for rate-coded representations and 1000 for our event-coded representations. Repulsion strength was initialized to unity. Local connectivity was set to 1Hz in estimating probability distances. We trained for 200 epochs at a learning rate initialized to unity. The resulting function was equipped with a learned graph of the data, which projected to the manifolds visualized in Figures 4, S14, and S17-18. An example of connections from such a learned graph were visualized (Figure 4F).

For our rate-coded manifold projections, the inclusion of predictive time bins ($p<0.05$) appeared sufficient for the separation of individuals (Figure S14A), which was supported by computing the minimum distance between the centroid of any individual and then comparing across multiple recording sessions. Minimum distances that were computed from predictive time bins exhibited a significantly smaller median value when compared to candidate time bins that were not predictive ($p>0.85$) according to a Wilcoxon-Mann-Whitney test ($p<0.001$, $N_{sessions}=29$), suggesting predictive activity leads to better separation of individuals in comparable rate-coded representations (Figure S14B). Shown are examples of rate-coded manifold projections that used predictive firing rates as trial-by-trial observations. Event-coded manifold projections used signed connection rates as spike-by-spike observations for Hades (Figure S12C,D) and for Baloo (Figure S14E,F) in addition to Archie (Figure S17A-C) and Buck (Figure S17D-F).

Estimating latent firing rate

Our latent firing rate was computed using unsupervised nonlinear dimensionality reduction of the absolute value of the signed connection rate for all neurons that had no less than one third of its computed signed connection rate values as positive (i.e. approximately symmetric). In computing the latent firing rate, we used a method of nonlinear dimensionality reduction that made no assumption of uniformity, which was achieved by passing the keyword argument, densmap=True to the manifold projection constructor, umap.UMAP, in the Python programming language. The output metric and the input metric were both Euclidean (flat), which supports the output having the same units as the input. The output was embedded in six-dimensional real space and the first three dimensions are visualized in Figure S16A for an exemplar recording session. After this output was computed at the spike times of all neurons involved, it was analyzed as a time series by time ordering the data according to evaluation time.

By considering latent firing rates evaluated at the times t=0 to 4 seconds after a stimulus onset, we observed relatively stable trajectories for multiple recording sessions conducted over multiple observers. Shown are three exemplar identity match trials, where Baloo observed the face and voice of her mother, her father, and her sister as shown in Figure S16B. We performed a median filter with a sliding window of 50 neuron spikes before plotting our estimates of the latent firing rates.

Generating the hammer bundle plot

The graph of connections bundled between individuals in Figure 4F represents the learned graph associated with an event-coded representation of identity analogous to Figure 4C. The procedure for generating the shape of Figure 4F was achieved using the Python function, umap.plot.connectivity with the keyword argument, edge_bundling='hammer'. Coloration was achieved to multiplying the resulting image with a color mask. The color mask resulted from passing the colored scatter plot of the event-coded representation through a Gaussian filter using the GNU Image Manipulation Program, which was also used for the image multiplication.

Determining anatomical positions of implants

All implants were followed by at least one postoperative MRI (Figure S20). The scans were aligned to anatomical features with RadiAnt Dicom viewer and the position along the anterior-posterior axis was determined by measurement from the center of the array to the ear canal. Because implants were stereotactically performed coronally, all recordings for a given array were assigned the same anterior-posterior (AP) position.

Because of the 1mm spread of the microwire brush arrays, it was difficult to precisely estimate the position of any given electrode, or indeed the entire bundle on a particular day. We used the position of the tip of the electrode from each MRI and extrapolated the trajectory by estimating position along the drive axis by cross-referencing with contemporaneous notes made of the date and distance of every movement of the drive. Based on a centroid at each estimated position, we chose particular sessions for we had the greatest confidence that the majority of the array was located predominantly in one or two hippocampal fields. Because the relative positioning of individual electrodes was not clearly observable, all reported analyses were developed to be agnostic to neuron location.

Confirming implant location by MRI

MRI was performed at the UCSD Center for Functional Magnetic Resonance Imaging in a 7.0T Bruker 20cm small animal imaging system using Advance II software. Preoperative images were analyzed in Osirix DICOM Viewer and stereotactic coordinates were established using a pair of saline-filled barrels affixed above the putative posterior end of temporal sulcus (marked on the skull during headcap surgery). Array positioning and tract trajectory was verified by post-operative MRI. Follow-up scans were performed occasionally to update array position.

Determination of anatomical positioning was performed using RadiAnt DICOM Viewer (Medixant, n.d.). Stereotactic alignment was performed using a number of clearly defined and readily identifiable anatomical landmarks. 2D coronal slices were made vertical by rotating to align the medial longitudinal fissure with a vertical line. Yaw was corrected by re-slicing the coronal plane to align both interaural canals. Pitch correction was performed by re-slicing MRI so that the 4^{th} ventricle was aligned vertically with the isthmus of the corpus callosum.

Position on the anterior-posterior axis was calculated relative to the interaural canal. Measurement was taken from the coronal slice at which the array first entered the hippocampal complex (Figure 1D, 2E). Arrays were implanted with as little pitch as possible, so AP position variability is negligible along the electrode trajectory.

Electrode positions are not precisely determinable with our brush arrays, as microwires are not visible at the resolution of the scans and individual tips are not individually distinguishable by any practical means available. Electrode splay of the 64-ch MBA in tissue was measured at approximately 1mm, so we approximated electrode position by use of a 1mm spherical voxel centered at the tip of the array.

We used a Microdrive with a 500µm thread pitch that could reliably make controlled movements with a precision of 30-40µm. An array tip was identified for every MRI in each subject and position was extrapolated based on contemporaneous notes regarding electrode movement. Once putative array centroids have been hand-tagged they were assigned to one of the hippocampal subfields. Centroids were deemed to be in a hippocampal subfield if more than 70% of their volume fell within that area, as assessed by hand-traced MRI. Recording sessions where the centroid fell significantly between two subregions were not counted in anatomical analyses. CA2 and CA3 were combined due to insufficient granularity in this methodology and resolution in our scans to effectively differentiate them. Figure S20 shows the estimated position of each electrode array in the hippocampus for all subjects.

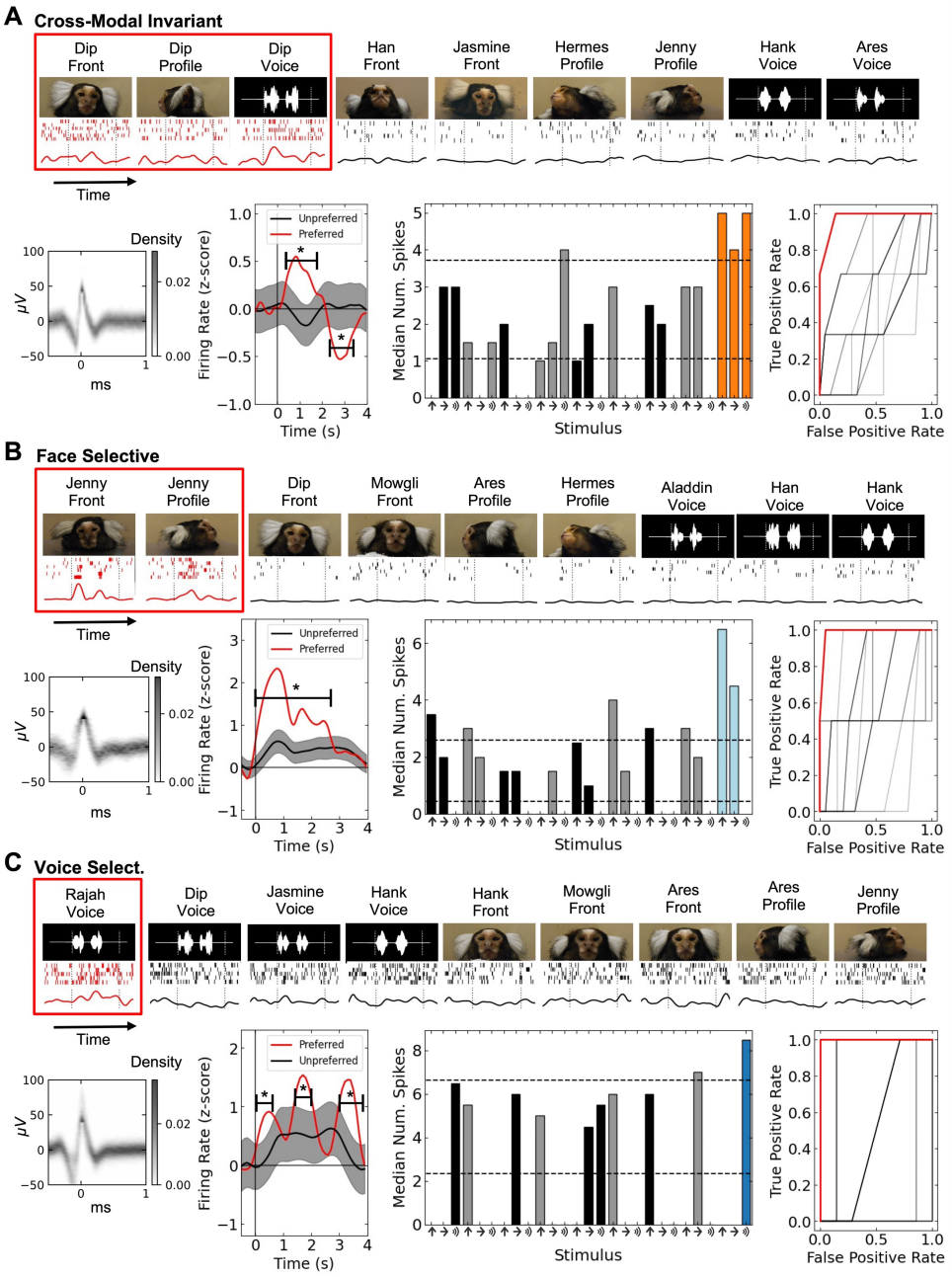

Fig. S1.

Supplementary Exemplar Neurons. Shown are exemplar identity neurons that are **(A)** cross-modal invariant**, (B)** face-selective, and **(C)** voice-selective comparable to Figure 1A-C. **(A-C)** Top row: subset of stimuli shown above raster and peristimulus time histogram (PSTH). Bottom row: spike waveform; normalized PSTH to all stimuli (preferred: red, nonpreferred: black), indicated are time points that show significant difference ($p<0.05$); median number of spikes for unimodal stimuli (grey/black indicate non-preferred individuals; ROC curve (shuffled controls shown in black). PSTH was normalized by the pre-stimulus baseline, and shaded regions indicate 95% confidence intervals. Indicated are time points that show a statistically significant difference in mean ($p<0.05$). Horizontal dotted lines indicate mean background firing rate and responsiveness threshold.

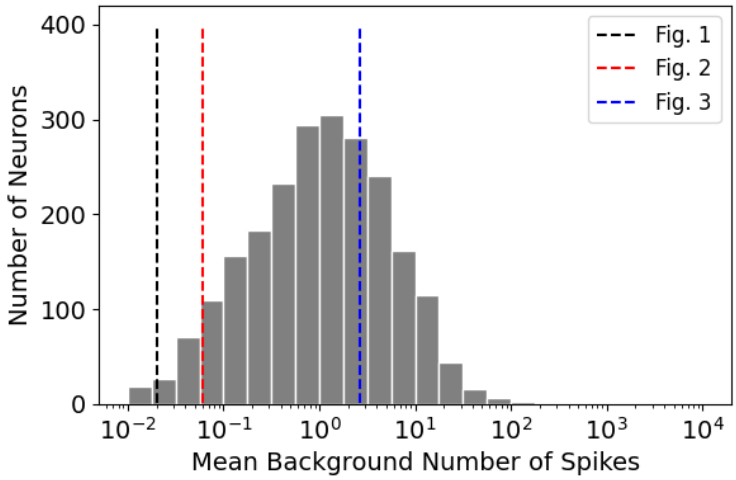

Fig. S2.
Marmoset hippocampus neurons have high baseline firing rates. Shown is a histogram of the mean background spike counts computed for all neurons involved in this study. The dotted lines come from the mean baselines reported in the main figures from Quian Quiroga *et al.*, *Nature* (2005), which summed over 700ms instead of 500ms. We confirm all recorded neurons are considered in this histogram.

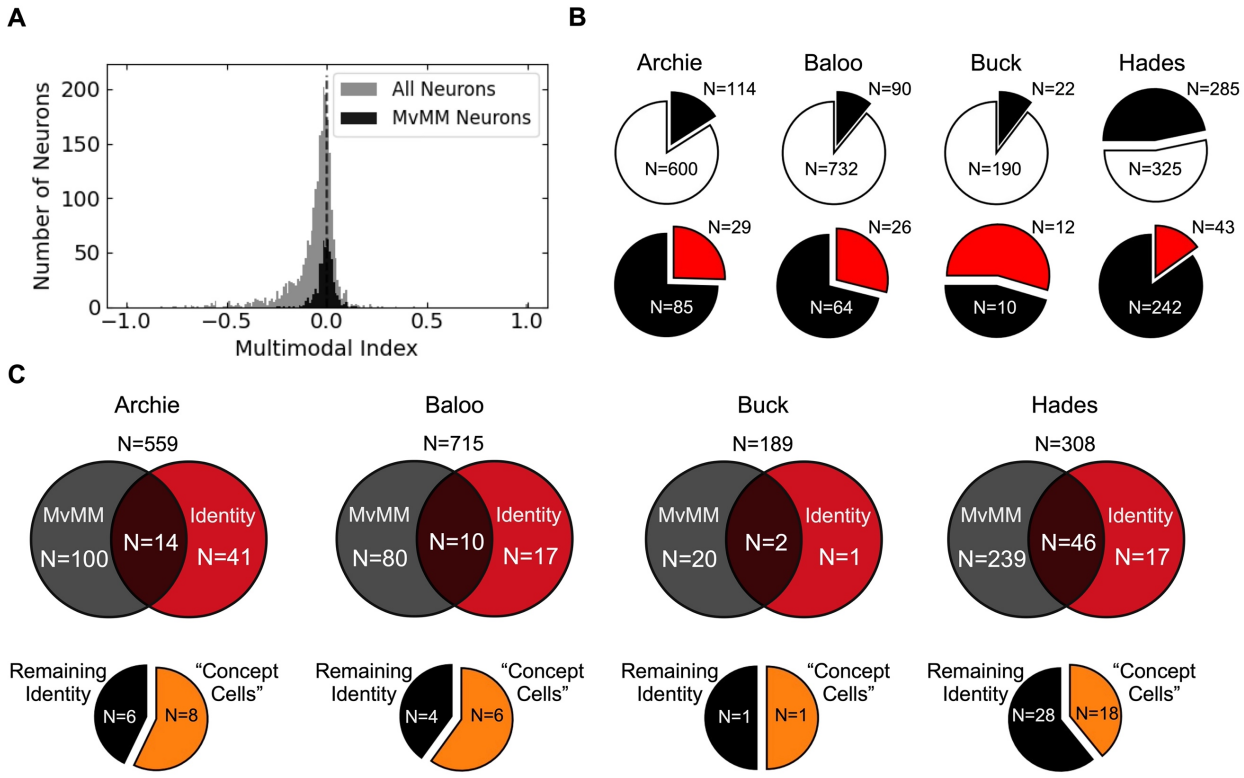

Fig. S3.

Distribution of neuron categories. (A) Histogram showing the multimodal index of MvMM neurons (black) and all recorded neurons (gray). Neither the mean nor median multimodal index was significantly greater than zero for either population ($p>0.05$, $N\geq499$). The multimodal index was not well defined for N=12 out of 511 MvMM neurons due to small response. Zero is indicated by the black dotted line. Bin width is 0.01. **(B)** Pie charts showing the abundance of MvMM neurons averaged over all recording sessions for each observer involved in this study. Shown is the number of MvMM neurons (black, top) amongst all other recorded neurons (white, top) and the number of identity-match preferring MvMM neurons (black, bottom) amongst the identity-mismatch preferring MvMM neurons (red, bottom). The CA1 region was only confirmed in $N_{sessions}=4$ out of 8 of the recording sessions from Hades. **(C)** Shown are (top) Venn diagrams and (bottom) pie charts that show the composition of populations investigated in the main text. (top) Venn diagram overlaps represent the abundance of (black) MvMM neurons in common with (red) identity-selective neurons, which exhibited a relative abundance of putative 'concept cells' as represented by the orange color in (bottom) the pie charts. Results are shown for each subject involved in this study. Furthermore, the number of MvMM neurons in common with MvMM predictive neurons was 359, while the number of MvMM neurons in common with the identity-specific predictive neurons was 388.

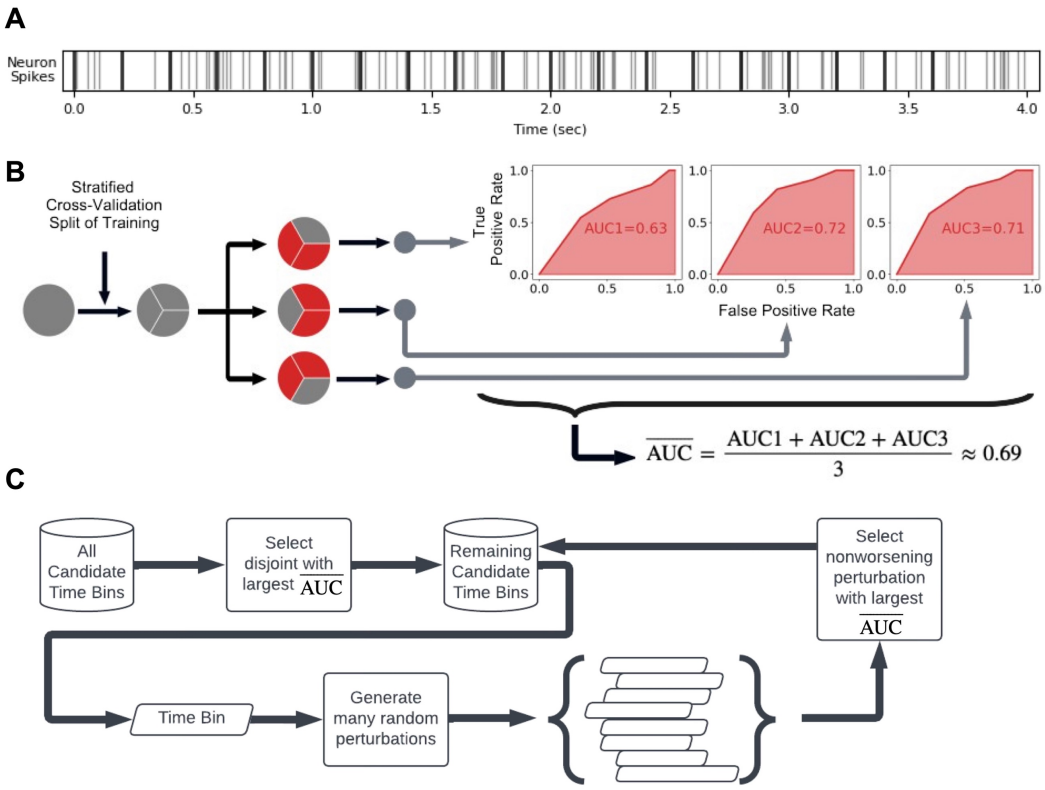

Fig. S4.
Identification of predictive time bins. (A) Schematic showing (gray) the spike times of an example neuron firing versus time after the stimulus onset at t=0. Indicated are (black) start and end times of time bins before the refinement procedure. **(B)** Flow chart showing training trials being split by stratified cross-validation to result in multiple ROC traces. Each training fold resulted in an area under the ROC curve (AUC), which were then averaged to produce the mean training AUC as an estimator of the general ability of a time bin to distinguish true trials from false trials. Time bins satisfying a list of properties were considered as candidate time bins (described in Methods). **(C)** Flow chart showing the procedure that resulted in all predictive time bins (described in Methods).

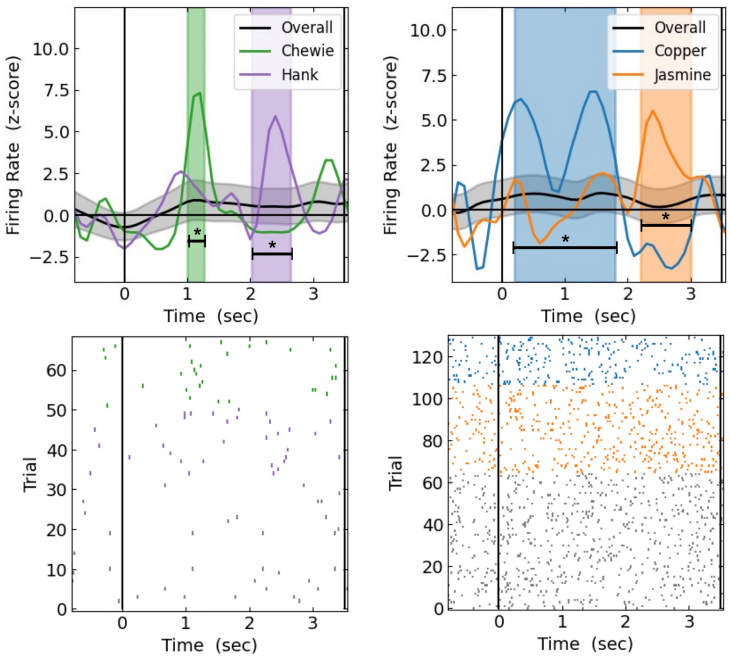

Fig. S5.

Exemplar predictive neurons. Shown are (top) PSTH traces and (bottom) spike rasters for two predictive neurons that each prefer at least two individuals. Shaded regions indicate predictive time bins, which exhibited a significantly different median firing rate for their preferred identity ($p<0.05$). Colors correspond to legends. The number of trials shown for Overall is matched in the plot to the number of trials for the two selective individuals.

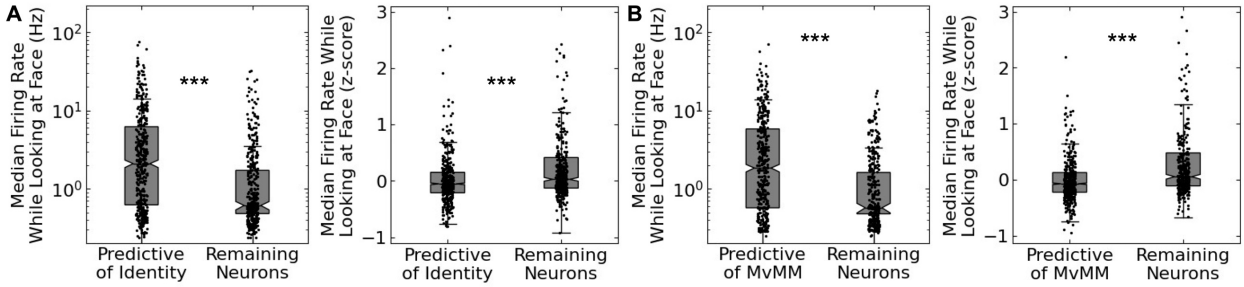

Fig. S6.

Predictive neurons driven by suppression of high background firing rate during face-viewing. (A-B) Boxplots showing median firing rate of (A) identity-specific and (B) MvMM predictive neurons *versus* the remaining neurons in terms of spikes per second (left) and z-score computed relative to baseline (right) averaged over times where the observer was looking at faces. Three asterisks indicate a significant difference in median according to a two-tailed Wilcox-Mann-Whitney test ($p<0.001$). The greater population was then confirmed by a one-tailed version of the test. We can infer that the predictive populations had relatively large baseline firing rates and that their mechanism may consist of the acquisition of inhibition or the release of inhibition. This mechanism explains the median firing rate being less than the remaining neurons (suppression) in terms of z-score, while the median firing rate is simultaneously larger for the predictive neurons relative to the remaining neurons.

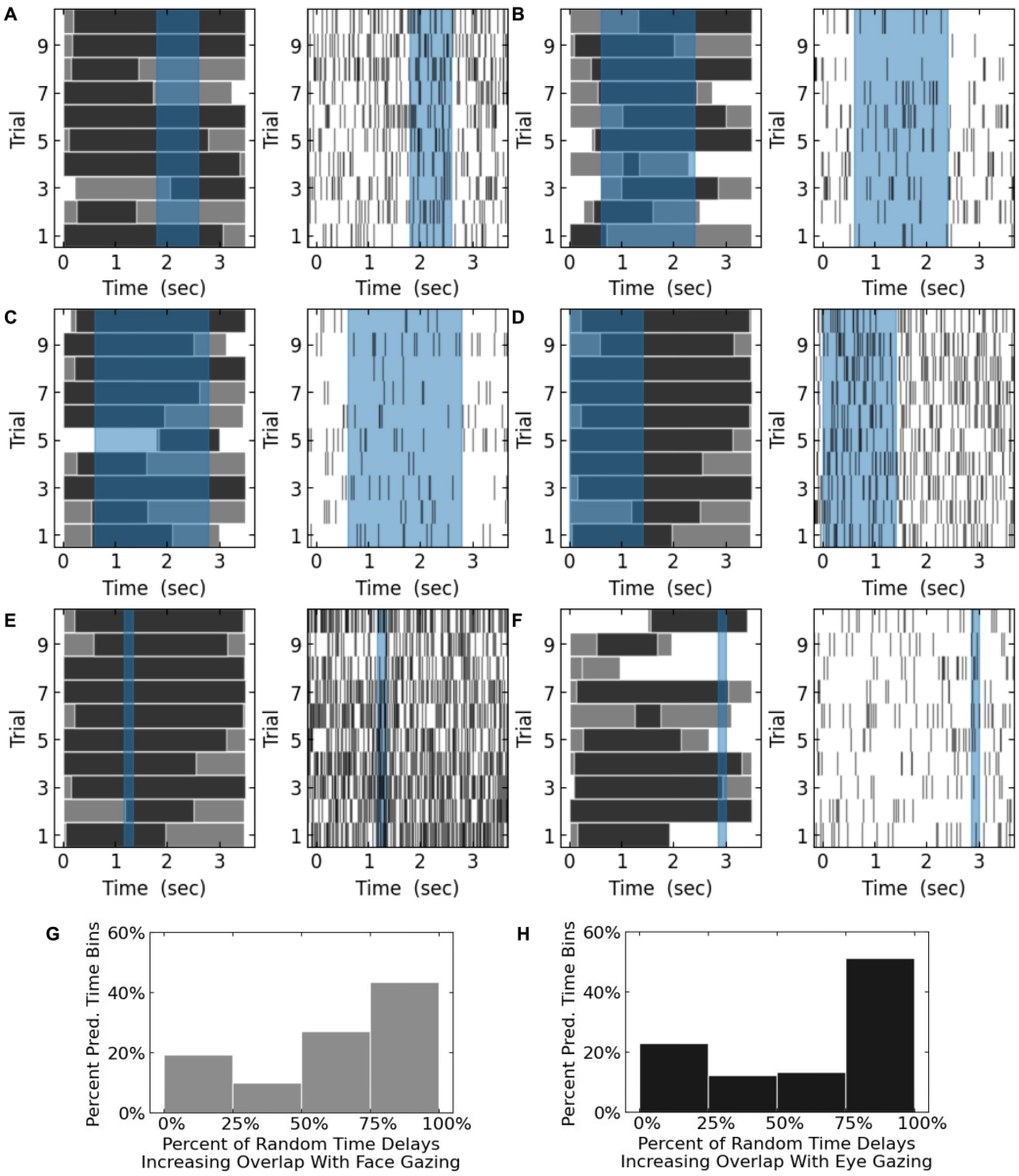

Fig. S7.

Variability of visual behavior relative to identity-specific predictive time bins. (A-F) Shown are visual behavior rasters (left) and spike rasters (right) for six predictive time bins. Blue shaded regions indicate the identity-specific predictive time bin. Gray indicates face gazing while black indicates eye gazing in the visual behavior rasters. Trials represent repeated presentations of the same front-facing unimodal stimulus. Unimodal stimuli were chosen to agree with the identity preference of the predictive time bin. **(G-H)** Histograms showing the relative abundance of random delays that increased the amount time in common between the time bin and time spent gazing at preferred faces (G, gray) and time spent gazing at preferred eyes (H, black). Bar height shows the percent of identity-specific predictive time bins, where each time bin had at least 10 presentations of at the same unimodal face-only stimulus where both eyes of the preferred individual were clearly visible (N_{bins}=218). More area in the right two bars indicates perturbing the time bins typically decreased overlap with visual behavior. Thus, more area in the right two bars relative to the left two bars supports the visual behavior being unrelated to predictive time bin occurrence.

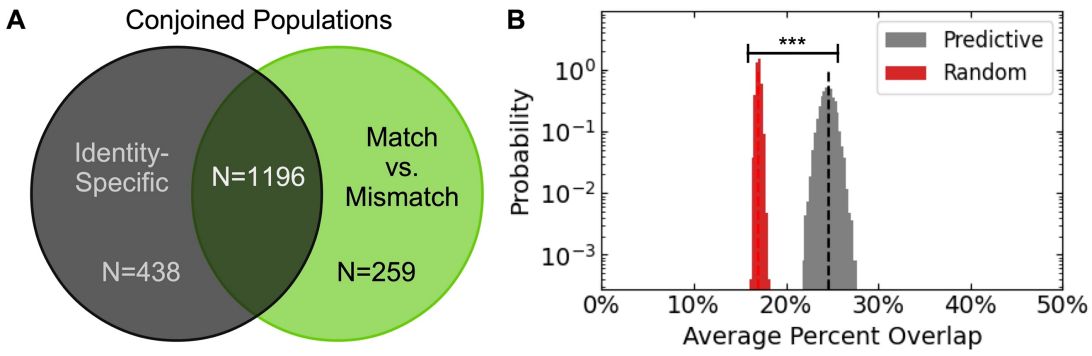

Fig. S8.

Overlap of predictive neuron populations. (A) Venn diagram showing the abundance of predictive neurons in common between the identity-specific predictive neurons (black) and the MvMM predictive neurons (green). **(B)** Histograms showing the probability density of the average percent overlap of the identity-specific predictive time bins with the MvMM predictive time bins from the same neurons (gray) and of an equal number of uniformly distributed pairs of random time bins as control (red). Indicated by the dashed lines is the total duration of overlap divided by the total duration of identity-specific predictive time bins, (612.3s/2493.7s)=24.6%±1.5% (black dashed line), which was significantly greater than control 17.0%±0.5% (red dashed line) according to Student's t-test ($p<0.001$, $N_{samples}=10{,}000$). Control uniformly sampled pairs of time bins on the interval from t=0 to t=3.5 seconds following stimulus onset. The bin width is 0.25%.

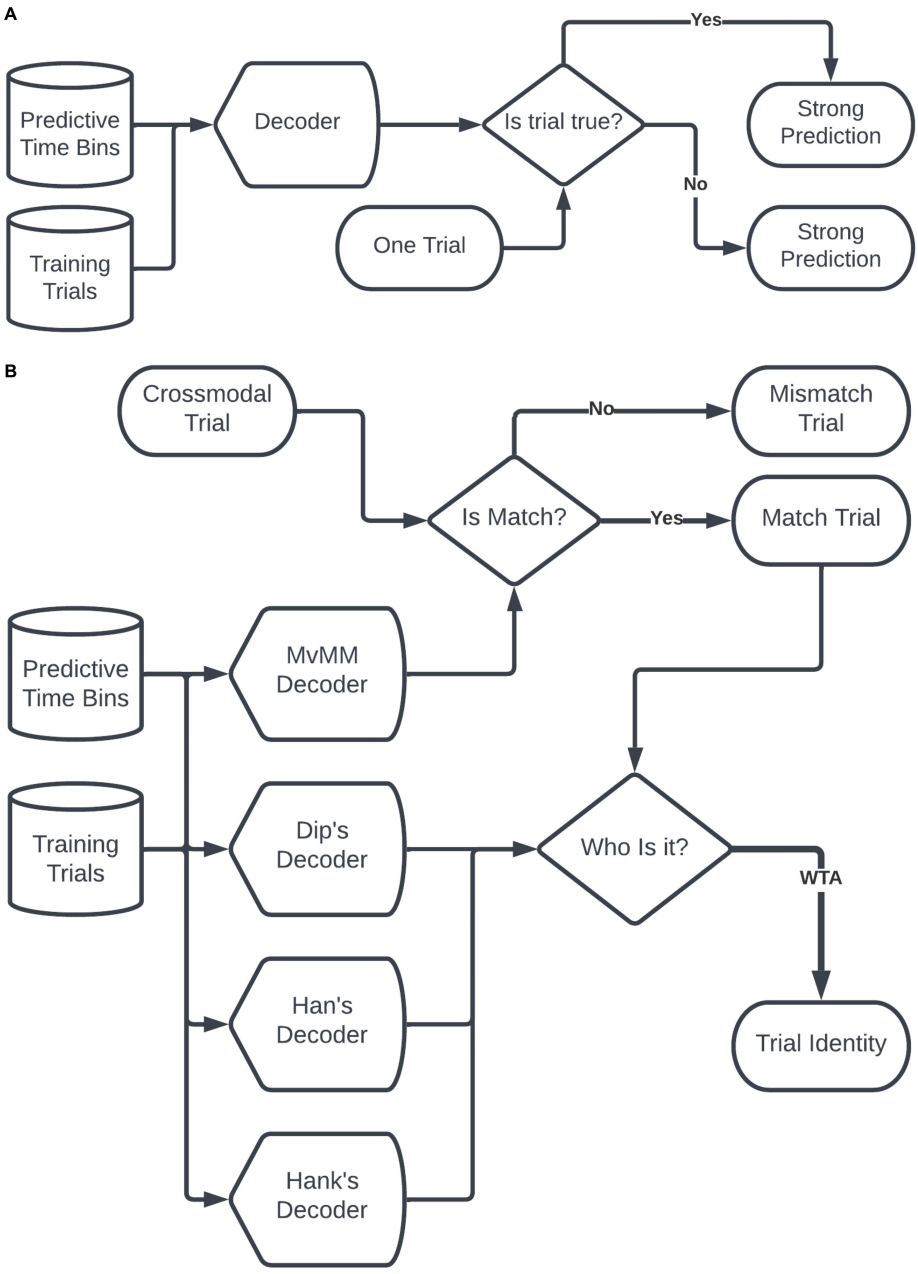

Fig. S9.
Decoder Schematics. (A) Flow chart showing predictive time bins were combined with the training trials that were used to determine the same predictive time bins to train a decoder for classifying trials as either true or false. The decoder then produced remarkably strong predictions on novel trials. **(B)** Flow chart showing the winner-take-all model resulting from a MvMM decoder and one identity-specific decoder for each individual. Cross-modal trials were categorized as either match or mismatch trials. The identity of the match trial was then predicted as that of the decoder with the largest output via winner-take-all (WTA).

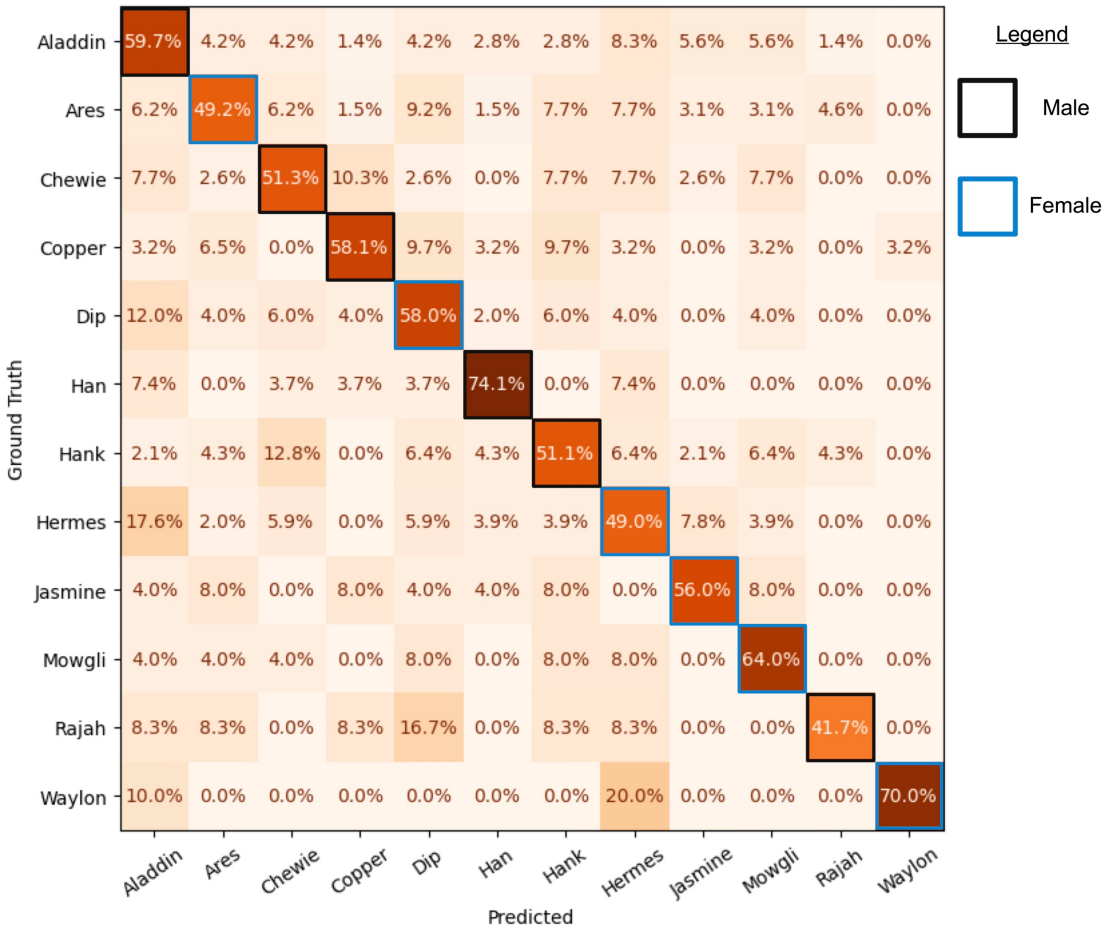

Fig. S10.
Multiple individuals classified by winner-take-all model. Confusion matrix reporting the winner-take-all predictions of the INM on twelve individuals shown to three observers over 34 recording sessions (testing accuracy=0.91, sensitivity=0.91, specificity=0.91, precision=0.88, negative predictive value=0.93, N_{trials}=454 match trials). The biological sex of the observed conspecifics is indicated by on the diagonal with blue indicating female and black indicating male. The following conspecifics were family members with a subject: Aladdin, Jasmine, Mowgli, Ares, Hermes. Percentages indicate true positive rates of the testing set of trials. All individuals decoded testing trials with a true positive rate at least 5X random chance, as is indicated by the black dashed line in Figure 3J of the main text.

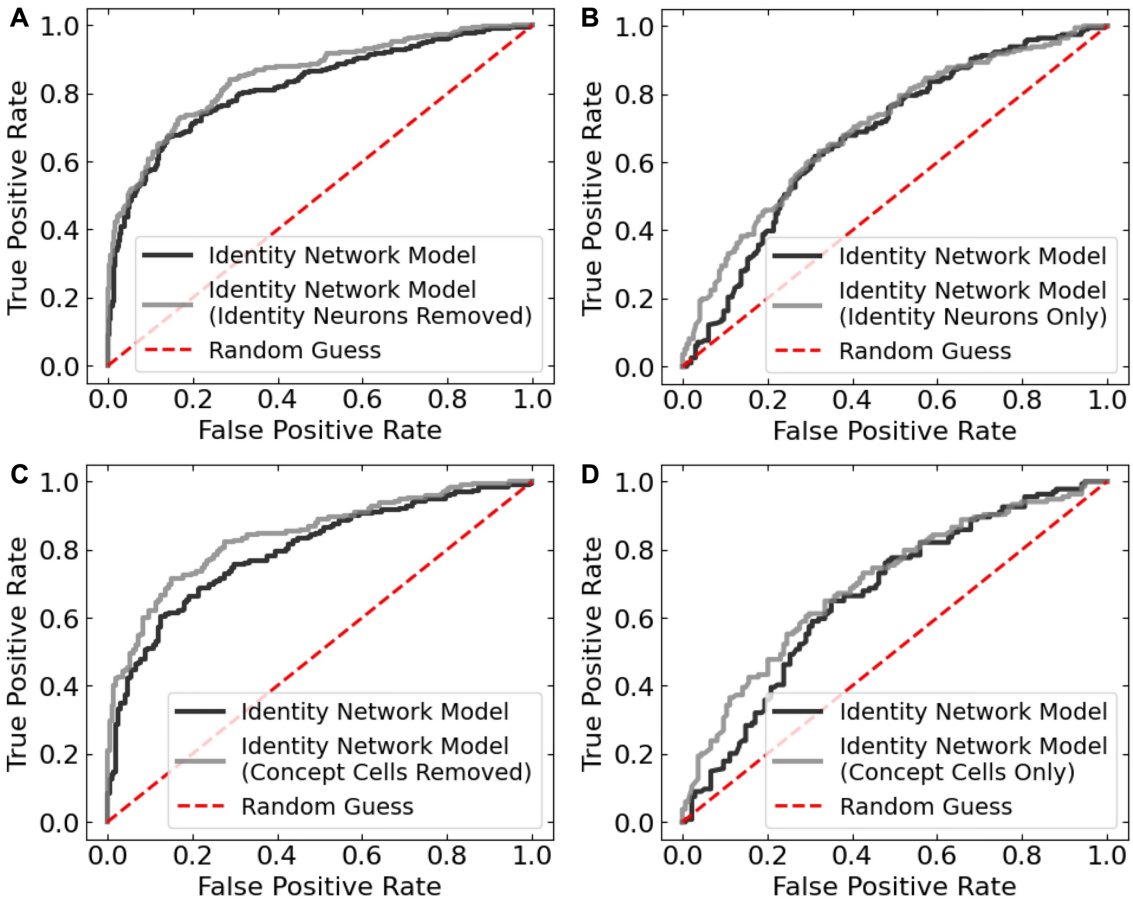

Fig. S11.
Decoding performance with and without identity selective neurons averaged over preferred individuals. (A) Shown are the ROC traces of the INM with all identity neurons removed (gray; AUC=0.850) and an equal number of random neurons removed from the remaining predictive population (black; AUC=0.820). **(B)** Shown are the ROC traces of the INM with only identity neurons considered (gray; AUC=0.700) and an equal number of neurons randomly selected from the remaining predictive population as control (black; AUC=0.677). Indicated is random chance (red dotted; AUC=0.500). **(C)** Shown are the ROC traces of the INM with all putative 'concept cells' removed (gray; AUC=0.841) and an equal number of random neurons removed from the remaining predictive population (black; AUC=0.795). **(D)** Shown are the ROC traces of the INM with only putative 'concept cells' considered (gray; AUC=0.700) and an equal number of neurons randomly selected from the remaining predictive population as control (black; AUC=0.666). Indicated is random chance (red dotted; AUC=0.500).

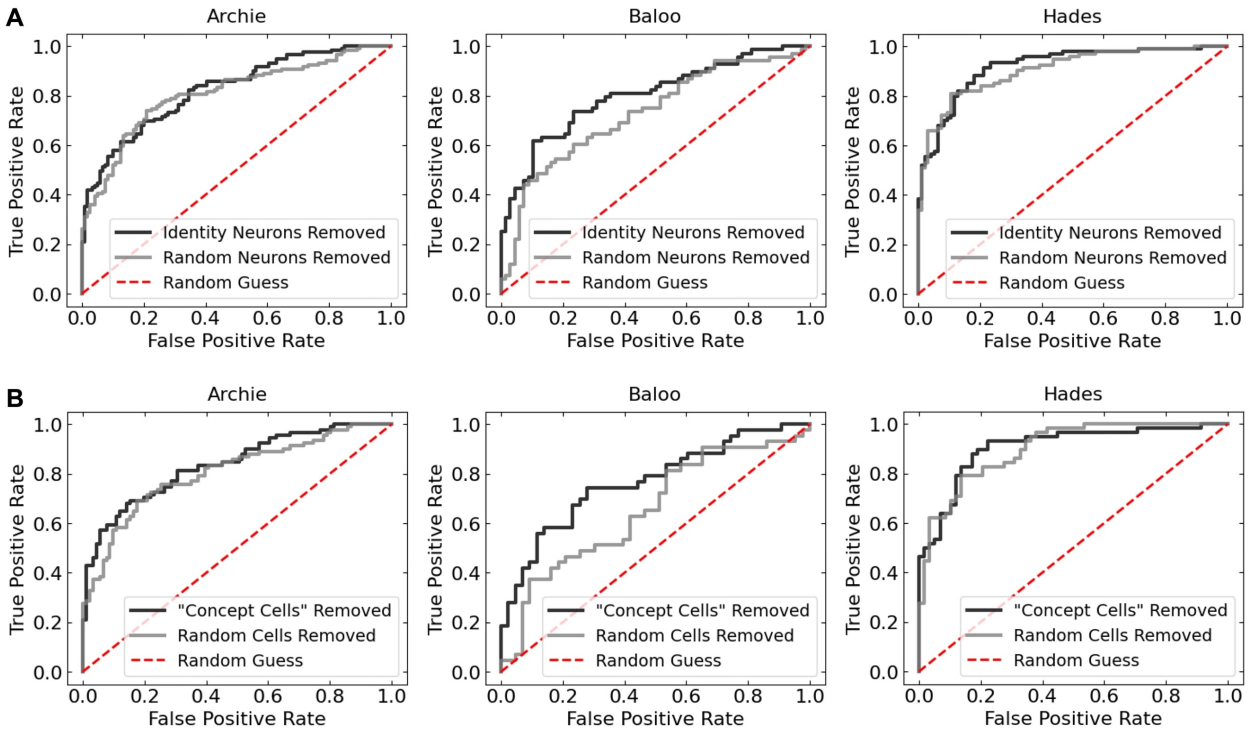

Fig. S12.

Identity network model for individual subjects. ROC curves were computed by averaging over all recording sessions for each of three observers: Archie (left, $N_{sessions}$=14), Baloo (middle, $N_{sessions}$=12), and Hades (right, $N_{sessions}$=8). **(A)** ROC curves of our INM with only identity neurons (black) and an equal number of cells from the remaining predictive population (gray). Individual identities were averaged over if they were preferred by at least one identity neuron. **(B)** ROC curves demonstrating the predictive power of our INM with all 'concept cells' removed (black) and an equal number of cells removed from the remaining predictive population (gray). Individual identities were averaged over if they were preferred by at least one 'concept cell'. We controlled for network size by using the same number of features for both ROC curves in each panel. We did this for both the MvMM predictive population and the identity-specific predictive population in evaluating the INM.

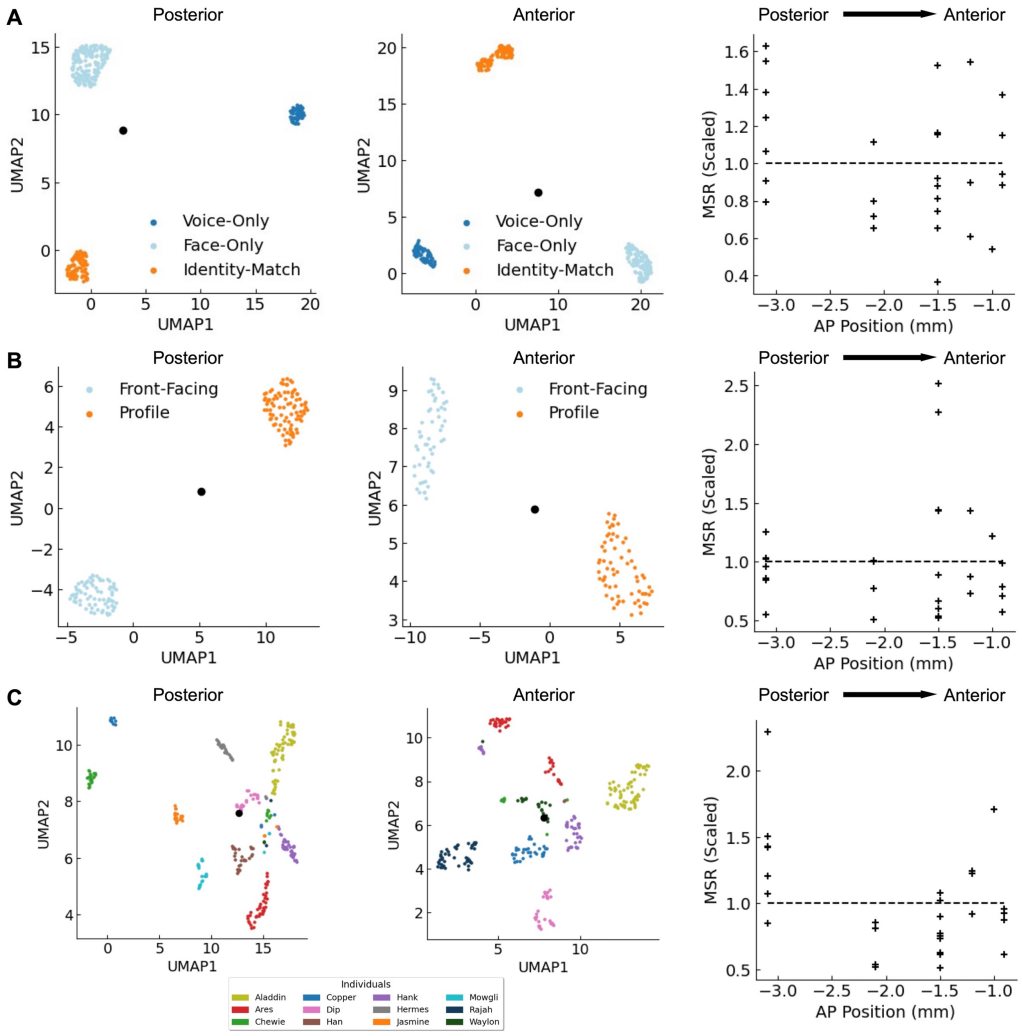

Fig. S13.

Separating stimulus categories at multiple probe locations. Shown are manifold projections from a single recording session conducted on (left) the most posterior and (middle) the most anterior probe location (anterior-posterior (AP) positions: -3.1mm, -0.9mm, respectively). (right) Shows AP positions *versus* mean-squared range (MSR) from (black dot in left and in middle) the mean projected trial location. MSR was scaled across recording sessions to have a mean value of unity, as is shown by the black dashed line in the scatter plot to the right. One symbol represents one recording session in the scatter plot to the right. Indicated is the direction from posterior to anterior hippocampus. **(A)** Shown are stimulus categories of mode (dark blue) voice-only trials, (light blue) face-only trials, and (orange) identity-match trials. **(B)** Shown are face-only trials categorized by orientation as either (light blue) front-facing or (orange) profile. **(C)** Shown are unimodal and identity match trials categorized by identity as is indicated by the legend. Large MSR values suggest excellent separation at the indicated probe location. In the majority of recording sessions conducted on the most posterior electrode array (AP position: -3.1 mm), MSR was greater than the mean, suggesting excellent separability of identity in the most posterior probe location. Input features were mean firing rates averaged over the stimulus from t=0 to 3.5 seconds for each recorded neuron. The most posterior probe at -3.1 mm was implanted in Hades, who generated all recording sessions confirmed to be in CA1. Recording sessions were omitted if their AP position was not confirmed to be the same within a 95% confidence of no more than ±0.1 mm, which resulted in 28 recording sessions being considered.

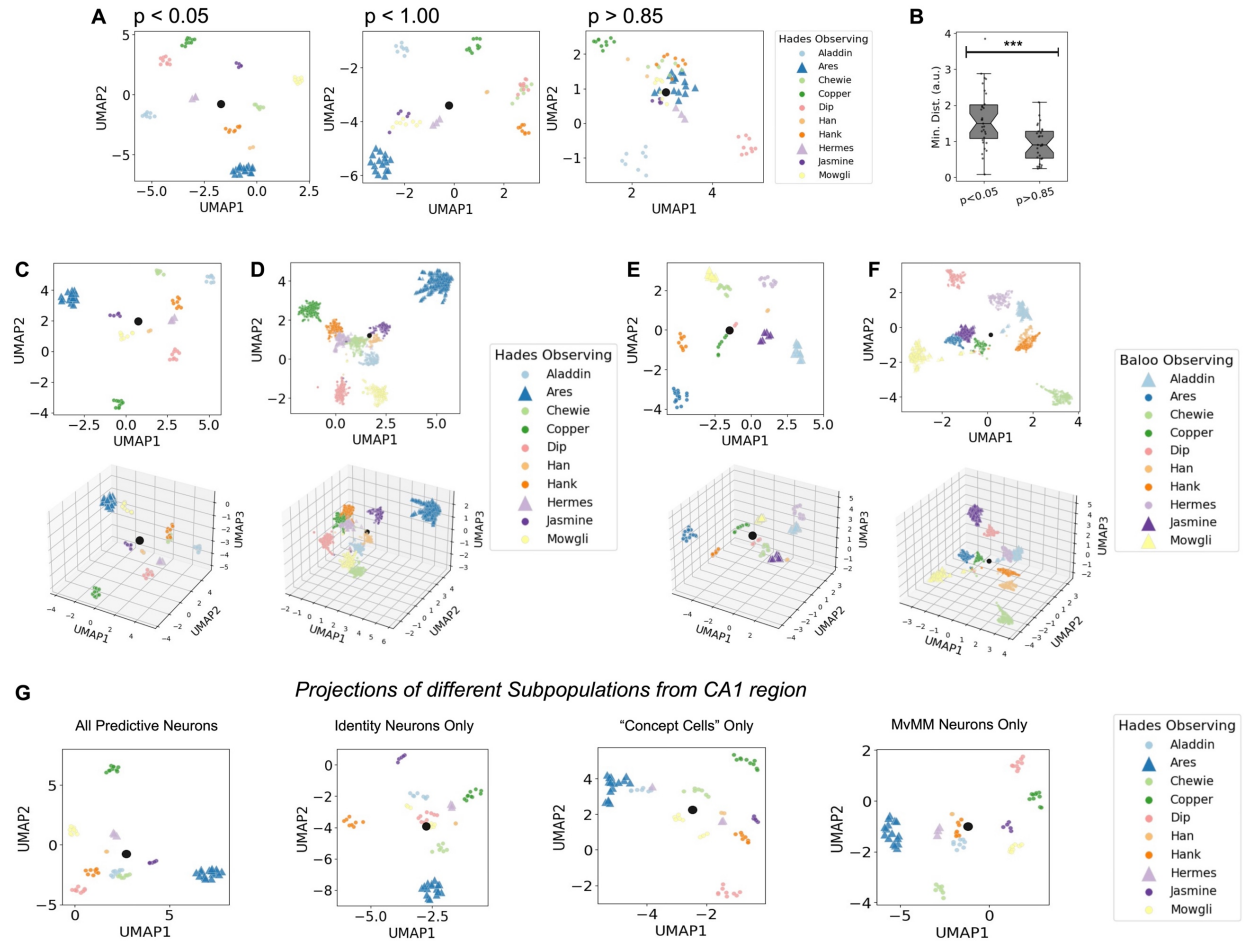

Fig. S14.

Low-dimensional projections of our rate code and event code. (A) Scatter plot showing an exemplar recording session as two-dimensional rate-coded representations of individual identity, where the firing rates were computed from all candidate time bins exhibiting (left) $p<0.05$, (middle) $p<1.00$, and (right) $p>0.85$. **(B)** Box-and-whisker plots showing the minimum distance between any individual in our rate-coded representation of individual identity. The median minimum distance of (left) $p<0.05$ was significantly smaller than the median minimum distance of (right) $p>0.85$ according to a Wilcoxon-Mann-Whitney test ($p<0.001$, $N_{sessions}=29$). **(C-F)** Shown are the (top) first two axes and (bottom) first three axes of our representations of individual identity for two distinct observers: (C,D) Hades and (E,F) Baloo. (C,E) Shown are manifold projections of our predictive time bins and (D,F) our signed connection rate. Colors indicate individuals, and triangles indicate family members. The signed connection rate was evaluated no more than two seconds after stimulus onset, which was evaluated whenever the neuron with the largest overall spike count fired. **(G)** Rate-coded manifold projections comparing the same recording session restricted to four subpopulations of identity-specific predictive neurons. Subpopulations are shown (from left to right): all identity-specific predictive neurons, all identity neurons, all cross-modal invariant 'concept cells', and all MvMM neurons. Colors indicate individual identities listed in legends. The recording session shown was confirmed to be in the CA1 region.

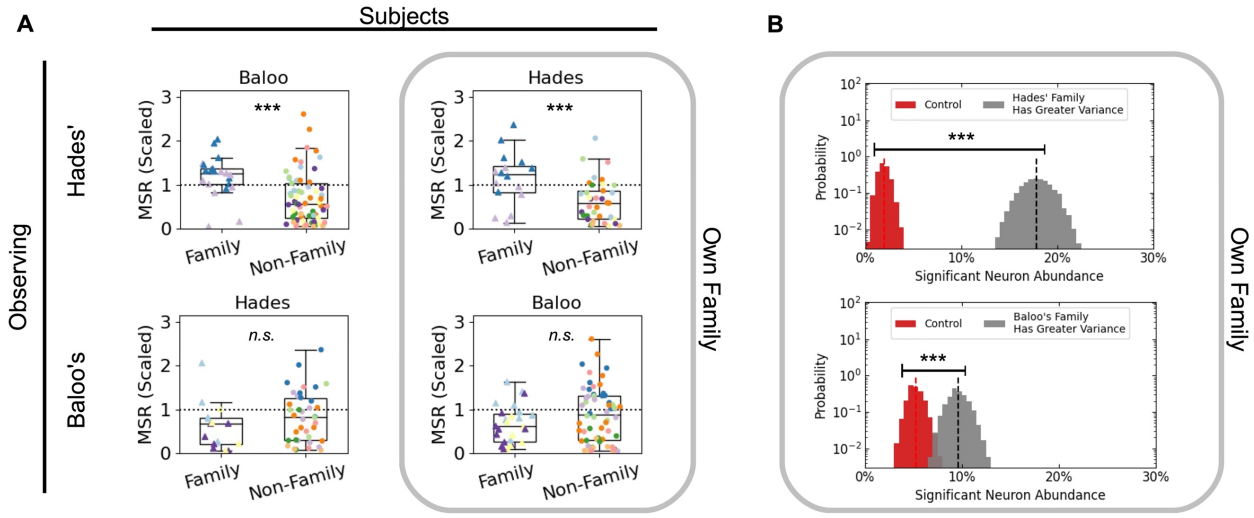

Fig. S15.

Separability of social categories. Some significantly different values when subjects were observing the family of (top) Hades and (bottom) Baloo. **(A)** Shown are boxplots of MSR of subjects observing families of other subjects. Significance was computed according to a one-sided Student's t-test consistent with the other subjects viewing the same family, resulting in (top left, $N_{identities} \geq 20$) $p<0.001$, (top right, $N_{identities} \geq 16$) $p<0.001$, (bottom left, $N_{identities} \geq 14$) $p=0.102$, and (bottom right, $N_{identities} \geq 26$) $p=0.055$. Gray box indicates subjects were observing their own families. **(B)** Histograms showing the relative abundance of neurons with significantly larger variance of signed connection rate for the subject's own family relative to other conspecifics according to Fligner-Killeen's test ($p<0.01$). Variance of signed connection rate was computed from the reference neuron to each neuron. Control was a random shuffle of the labels for each neuron. Distributions were determined via bootstrap. Dotted lines indicate the mean values for Hades viewing her own family (left, 18±3% out of N=610) and Baloo viewing her own family (right, 10±2% out of N=822), which both exhibited significantly more significant neurons than control (left, 2.0±1.1% out of N=610; right, 5.2±1.5% out of N=822) according to Student's t-test ($p<0.001$, $N_{bootstrap}=10{,}000$). Uncertainty indicates 95% confidence intervals of the mean. Gray box indicates subjects were observing their own families. Bin width is 0.5%.

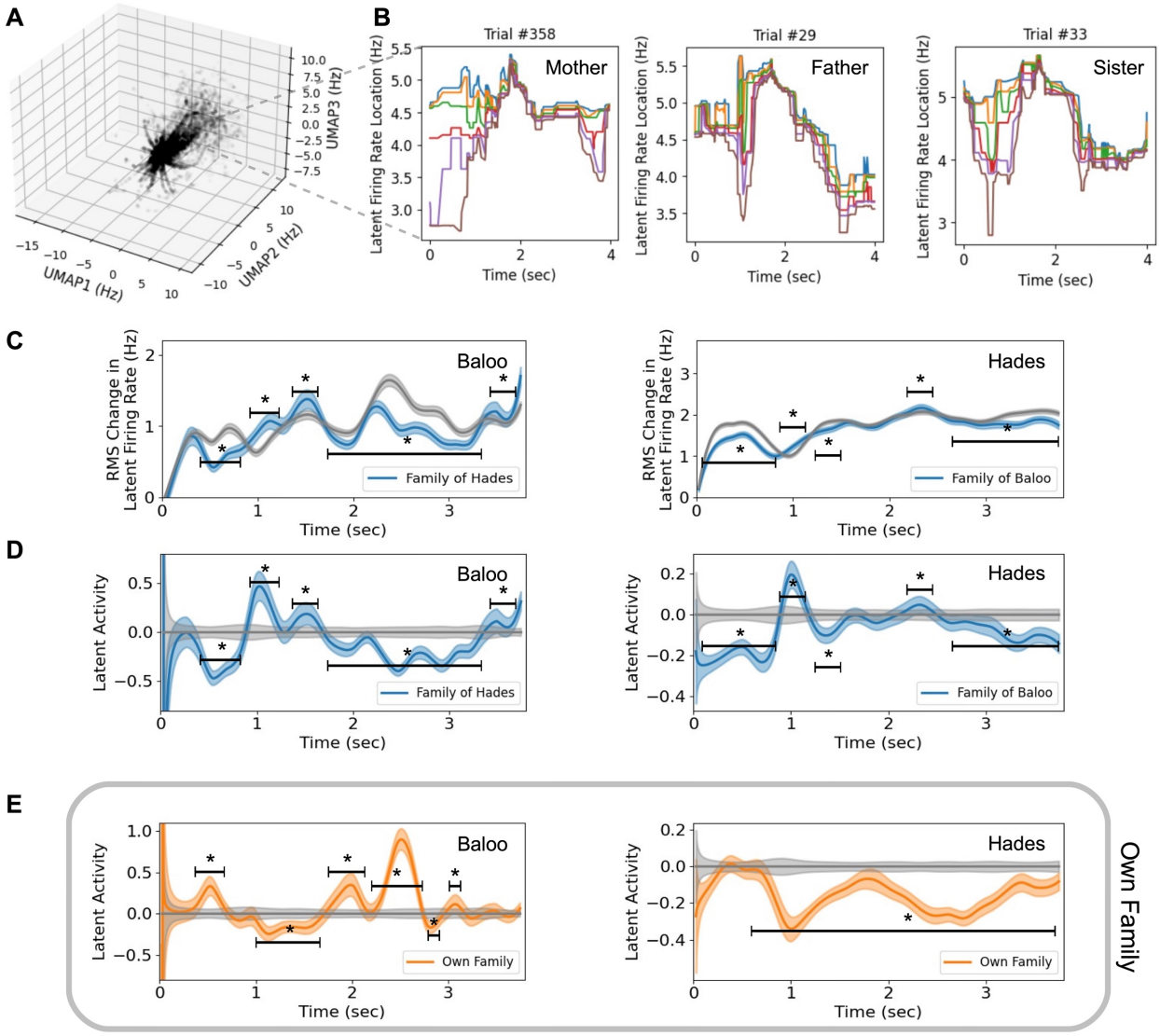

Fig. S16.

Quantification of latent activity. (A) Shown are the first three axes of the six-dimensional latent firing rate, which was an unsupervised manifold projection of the absolute value of the signed connection rate from the same neuron with the largest overall spike count (i.e. the reference neuron) to all neurons that appeared approximately symmetric (defined in Methods). **(B)** Shown are time traces of our latent firing rate for an exemplary trial from each of three family members of Baloo. Each color represents one dimension. The color of dimensions is consistent between panels. **(C)** Root mean squared (RMS) change in latent firing rate *versus* time averaged over all recording sessions from subjects (left) Baloo and (right) Hades. Traces average over identity match trials showing (blue) the family members of the subject and (gray) all conspecifics. **(D)** Latent activity *versus* time for (left) Baloo and (right) Hades. Latent activity traces were computed as the ratio of the RMS change in latent firing rate to control minus one. Control was RMS change in latent firing rate averaged over all identity match trials. **(E)** Latent activity *versus* time for (left) Baloo and (right) Hades viewing their own family. Control was as in (D).

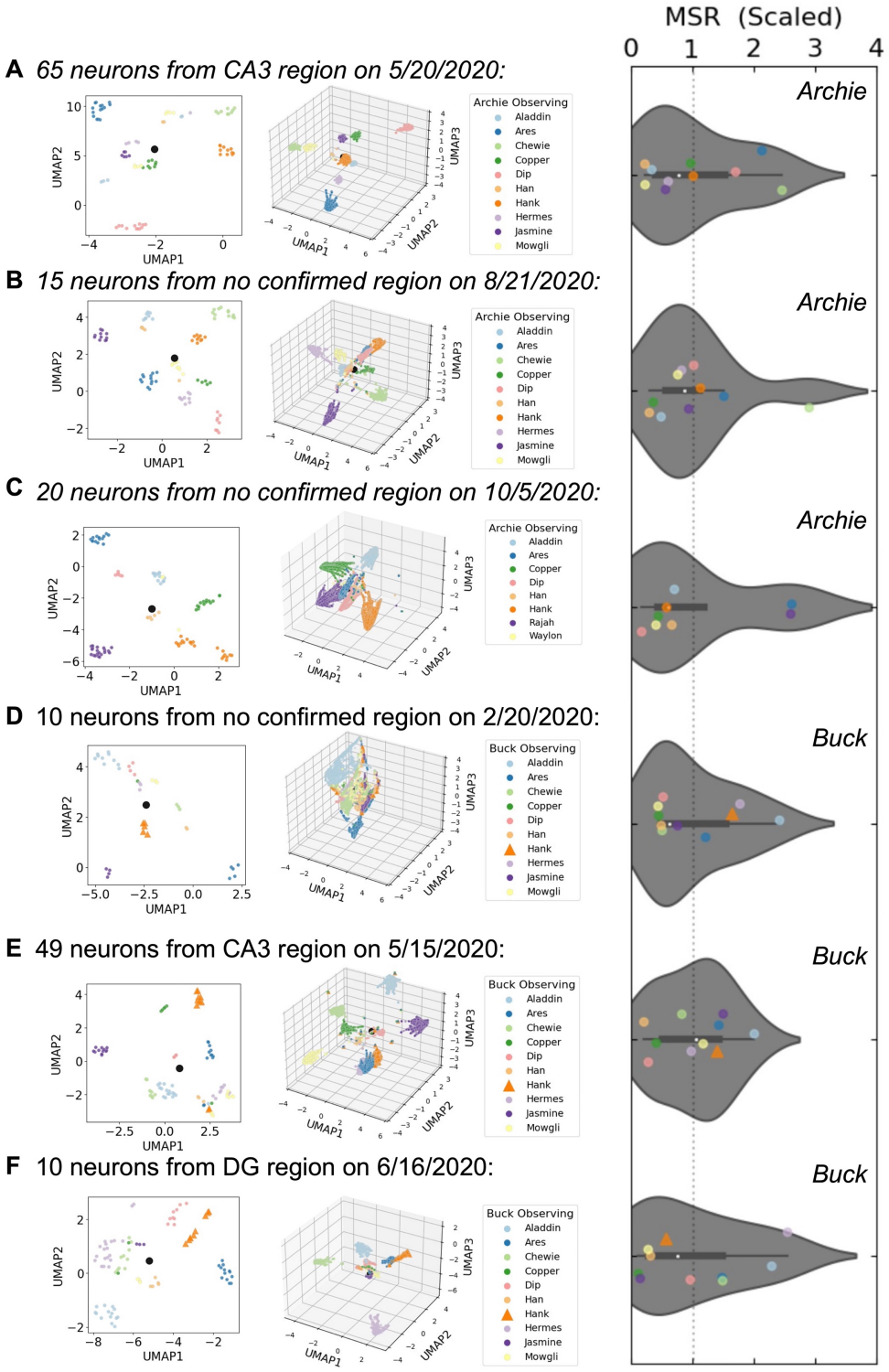

Fig. S17.
Separability of individuals over long time scales. (A-F) Manifold projections comparing three different recording sessions conducted on different observers, Archie (A-C) and Buck (D-F). Shown are rate-coded projections (left), event-coded projections (middle), and MSR computed from the event-coded projections (right). Colors indicate individual identities listed in legends. Triangles indicate family members.

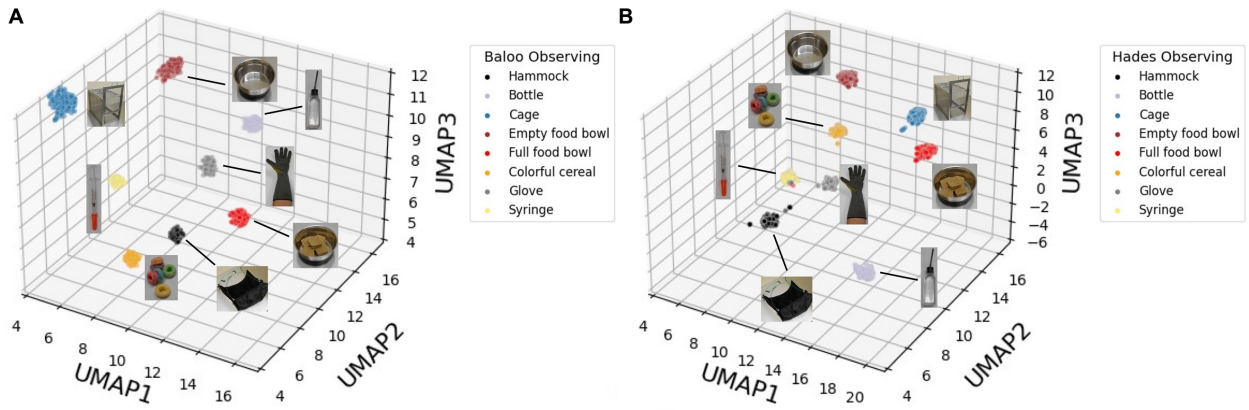

Fig. S18.
Separability of socially-agnostic categories. (A-B) Event-coded representations of inanimate objects from the laboratory setting. Separation of socially-agnostic categories are shown in marmoset hippocampus for two subjects, (A) Baloo and (B) Hades. Visual images from each of these object categories was presented to subjects using the same stimulus presentation protocol as for the unimodal stimuli while recording single neuron activity in marmoset hippocampus from two marmosets. Likewise, we performed the same signed-connection rate analysis and input those data into UMAP using the same data analysis pipeline as described for analyses presented in Figure 4.

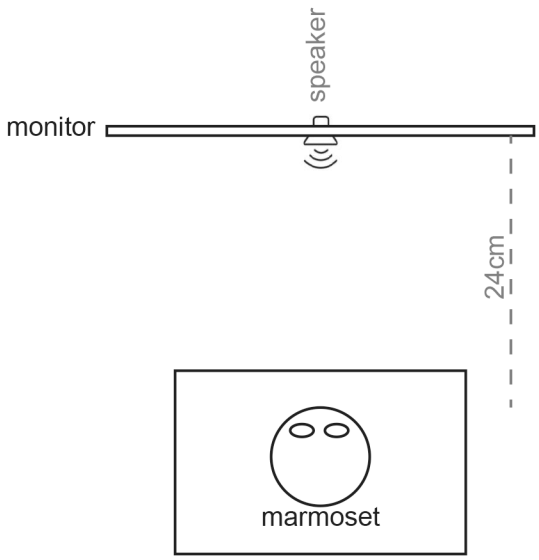

Fig. S19.
Schematic drawing of experimental setup. In an anechoic chamber, marmoset subjects were seated, positioned 24 centimeters away from a monitor and a speaker. The speaker was located just below the monitor.

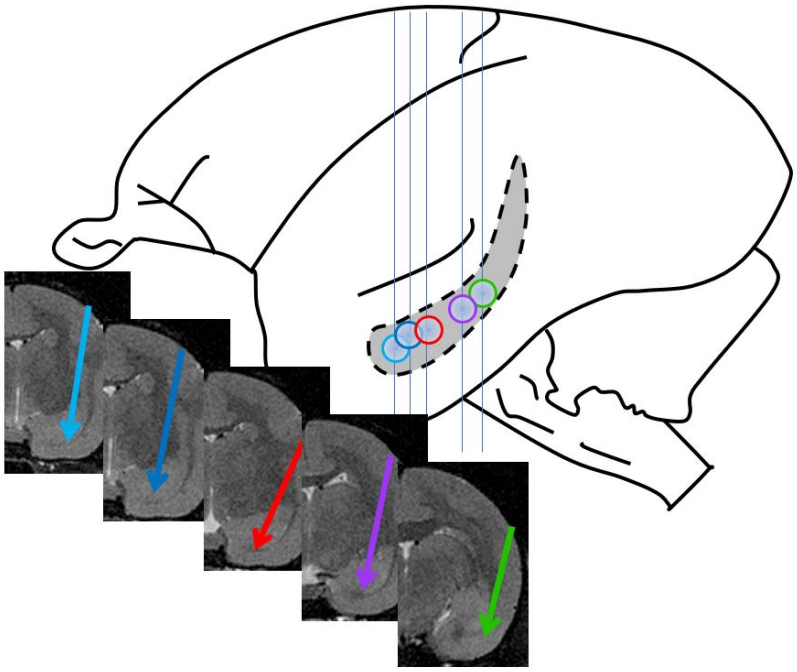

Fig. S20.
Anatomical locations of microwire bundles across animals. Arrows on MRI cross-sections indicate trajectory of each microwire brush array in marmoset hippocampus. Each color indicates a different animal's array. Circles correspond to anterior-posterior position.

Hyperparameter	MvMM	identity-specific
base_score	0.5	0.5
num_parallel_tree	25	50
subsample	0.2	0.2
colsample_bynode	0.1	0.1
max_delta_step	0.5	1
reg_alpha	0.4	0.3
reg_lambda	0.4	0.3
gamma	0.1	5
max_depth	5	2
learning_rate	0.9	0.6
min_child_weight	0.5	1

Table S1 Table of hyperparameters for our population-level neural decoders. Numerical values were passed as keyword arguments to the constructor of xgboost.XGBClassifier instances *(21)*. Columns correspond to the two types of predictive populations reported in the main text.

References and Notes

1. R. M. Seyfarth, D. L. Cheney, in *The Evolution of Primate Societies*, J. Mitani, J. Call, P. M. Kappeler, R. Palombit, J. B. Silk, Eds. (Univ. Chicago Press, 2012), pp. 629–642.

2. C. Perrodin, C. Kayser, N. K. Logothetis, C. I. Petkov, Voice cells in the primate temporal lobe. *Curr. Biol.* **21**, 1408–1415 (2011). doi:10.1016/j.cub.2011.07.028 Medline

3. P. Belin, C. Bodin, V. Aglieri, A "voice patch" system in the primate brain for processing vocal information? *Hear. Res.* **366**, 65–74 (2018). doi:10.1016/j.heares.2018.04.010 Medline

4. J. Sliwa, A. Planté, J.-R. Duhamel, S. Wirth, Independent Neuronal Representation of Facial and Vocal Identity in the Monkey Hippocampus and Inferotemporal Cortex. *Cereb. Cortex* **26**, 950–966 (2016). doi:10.1093/cercor/bhu257 Medline

5. U. Rutishauser, O. Tudusciuc, D. Neumann, A. N. Mamelak, A. C. Heller, I. B. Ross, L. Philpott, W. W. Sutherling, R. Adolphs, Single-unit responses selective for whole faces in the human amygdala. *Curr. Biol.* **21**, 1654–1660 (2011). doi:10.1016/j.cub.2011.08.035 Medline

6. M. C. Rose, B. Styr, T. A. Schmid, J. E. Elie, M. M. Yartsev, Cortical representation of group social communication in bats. *Science* **374**, eaba9584 (2021). doi:10.1126/science.aba9584 Medline

7. L. Chang, D. Y. Tsao, The Code for Facial Identity in the Primate Brain. *Cell* **169**, 1013–1028.e14 (2017). doi:10.1016/j.cell.2017.05.011 Medline

8. R. Báez-Mendoza, E. P. Mastrobattista, A. J. Wang, Z. M. Williams, Social agent identity cells in the prefrontal cortex of interacting groups of primates. *Science* **374**, eabb4149 (2021). doi:10.1126/science.abb4149 Medline

9. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005). doi:10.1038/nature03687 Medline

10. R. Quian Quiroga, A. Kraskov, C. Koch, I. Fried, Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr. Biol.* **19**, 1308–1313 (2009). doi:10.1016/j.cub.2009.06.060 Medline

11. I. V. Viskontas, R. Q. Quiroga, I. Fried, Human medial temporal lobe neurons respond preferentially to personally relevant images. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21329–21334 (2009). doi:10.1073/pnas.0902319106 Medline

12. R. Q. Quiroga, Concept cells: The building blocks of declarative memory functions. *Nat. Rev. Neurosci.* **13**, 587–597 (2012). doi:10.1038/nrn3251 Medline

13. R. Quian Quiroga, An integrative view of human hippocampal function: Differences with other species and capacity considerations. *Hippocampus* **33**, 616–634 (2023). doi:10.1002/hipo.23527 Medline

14. H. S. Courellis, S. U. Nummela, M. Metke, G. W. Diehl, R. Bussell, G. Cauwenberghs, C. T. Miller, Spatial encoding in primate hippocampus during free navigation. *PLOS Biol.* **17**, e3000546 (2019). doi:10.1371/journal.pbio.3000546 Medline

15. F. Mormann, S. Kornblith, R. Q. Quiroga, A. Kraskov, M. Cerf, I. Fried, C. Koch, Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J. Neurosci.* **28**, 8865–8872 (2008). doi:10.1523/JNEUROSCI.1640-08.2008 Medline

16. J. Minxha, C. Mosher, J. K. Morrow, A. N. Mamelak, R. Adolphs, K. M. Gothard, U. Rutishauser, Fixations gate species-specific responses to free viewing of faces in the human and macaque amygdala. *Cell Rep.* **18**, 878–891 (2017). doi:10.1016/j.celrep.2016.12.083 Medline

17. M. Rigotti, O. Barak, M. R. Warden, X.-J. Wang, N. D. Daw, E. K. Miller, S. Fusi, The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013). doi:10.1038/nature12160 Medline

18. C. T. Miller, D. Gire, K. Hoke, A. C. Huk, D. Kelley, D. A. Leopold, M. C. Smear, F. Theunissen, M. Yartsev, C. M. Niell, Natural behavior is the language of the brain. *Curr. Biol.* **32**, R482–R493 (2022). doi:10.1016/j.cub.2022.03.031 Medline

19. D. Kumaran, E. A. Maguire, Match mismatch processes underlie human hippocampal responses to associative novelty. *J. Neurosci.* **27**, 8517–8524 (2007). doi:10.1523/JNEUROSCI.1677-07.2007 Medline

20. M. Fyhn, S. Molden, S. Hollup, M.-B. Moser, E. Moser, Hippocampal neurons responding to first-time dislocation of a target object. *Neuron* **35**, 555–566 (2002). doi:10.1016/S0896-6273(02)00784-5 Medline

21. T. Chen, C. Guestrin, "XGBoost: A Scalable Tree Boosting System" in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, 2016), pp. 785–794.

22. E. H. Nieh, M. Schottdorf, N. W. Freeman, R. J. Low, S. Lewallen, S. A. Koay, L. Pinto, J. L. Gauthier, C. D. Brody, D. W. Tank, Geometry of abstract learned knowledge in the hippocampus. *Nature* **595**, 80–84 (2021). doi:10.1038/s41586-021-03652-7 Medline

23. W. A. Freiwald, D. Y. Tsao, Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330**, 845–851 (2010). doi:10.1126/science.1194908 Medline

24. T. P. Reber, M. Bausch, S. Mackay, J. Boström, C. E. Elger, F. Mormann, Representation of abstract semantic knowledge in populations of human single neurons in the medial temporal lobe. *PLOS Biol.* **17**, e3000290 (2019). doi:10.1371/journal.pbio.3000290 Medline

25. P. Baraduc, J.-R. Duhamel, S. Wirth, Schema cells in the macaque hippocampus. *Science* **363**, 635–639 (2019). doi:10.1126/science.aav5404 Medline

26. J. Sliwa, J.-R. Duhamel, O. Pascalis, S. Wirth, Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1735–1740 (2011). doi:10.1073/pnas.1008169108 Medline

27. I. Adachi, R. R. Hampton, Rhesus monkeys see who they hear: Spontaneous cross-modal memory for familiar conspecifics. *PLOS ONE* **6**, e23345 (2011). doi:10.1371/journal.pone.0023345 Medline

28. D. Y. Tsao, M. S. Livingstone, Mechanisms of face perception. *Annu. Rev. Neurosci.* **31**, 411–437 (2008). doi:10.1146/annurev.neuro.30.051606.094238 Medline

29. K. M. Gothard, F. P. Battaglia, C. A. Erickson, K. M. Spitler, D. G. Amaral, Neural responses to facial expression and face identity in the monkey amygdala. *J. Neurophysiol.* **97**, 1671–1683 (2007). doi:10.1152/jn.00714.2006 Medline

30. S. M. Landi, P. Viswanathan, S. Serene, W. A. Freiwald, A fast link between face perception and memory in the temporal pole. *Science* **373**, 581–585 (2021). doi:10.1126/science.abi6671 Medline

31. J. Munuera, M. Rigotti, C. D. Salzman, Shared neural coding for social hierarchy and reward value in primate amygdala. *Nat. Neurosci.* **21**, 415–423 (2018). doi:10.1038/s41593-018-0082-8 Medline

32. W. A. Freiwald, Social interaction networks in the primate brain. *Curr. Opin. Neurobiol.* **65**, 49–58 (2020). doi:10.1016/j.conb.2020.08.012 Medline

33. T. Tyree, M. Metke, C. Miller, Cross-modal representation of identity in primate hippocampus, Dataset, Dryad (2022); https://doi.org/10.5061/dryad.qnk98sfkv.

34. J. F. Mitchell, J. H. Reynolds, C. T. Miller, Active vision in marmosets: A model system for visual neuroscience. *J. Neurosci.* **34**, 1183–1194 (2014). doi:10.1523/JNEUROSCI.3899-13.2014 Medline

35. M. J. Jutras, E. A. Buffalo, Recognition memory signals in the macaque hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 401–406 (2010). doi:10.1073/pnas.0908378107 Medline

36. C. T. Miller, A. Wren Thomas, Individual recognition during bouts of antiphonal calling in common marmosets. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **198**, 337–346 (2012). doi:10.1007/s00359-012-0712-7 Medline

37. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree" in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, Eds. (Curran Associates, 2017).

38. L. McInnes, J. Healy, N. Saul, L. Großberger, UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018). doi:10.21105/joss.00861