A multispecies perspective on the evolution of form vision

Thesis by

Frank F. Lanfranchi

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2025

(Defended Dec. 19th, 2024)

Frank F. Lanfranchi
ORCID: 0000-0001-8176-320X

DEDICATION

Ai miei genitori,
Per l'incommensurabile dono
di un'educazione

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Doris Tsao, for the incredible opportunity she gave me. When I arrived at Caltech as a graduate student, viewing the brain as nothing more than a 1.3 kg ball of tissue, she handed me a box of electrodes and encouraged me to listen to the neurons. I have been listening ever since. Doris embodies a life devoted to scientific discovery, a life that brings to mind a quote from Dante's Inferno, which in translation reads: "You were not made to live as brutes, but to follow virtue and knowledge." She has inspired me to think deeply, pursue experiments with rigor and curiosity, and never shy away from the big questions.

I would like to extend my thanks to my committee members: Carlos Lois for his constant support and guidance through the uncertainties and frustrations of graduate school; Mikhail Shapiro for his generosity and his willingness to share his scientific discoveries in functional ultrasound with me and my lab; Richard Andersen for always challenging me both in class and beyond with his thought-provoking questions; and to my newly acquired members, Ueli Rutishauser and Pietro Perona, for their patience and generosity with their time. I owe deep thanks to Daniel Wagenaar for his friendship and constant encouragement, which have been crucial to my success in graduate school. Daniel taught me the value of rigorous scientific methods and hypothesis testing, and showed me that complex systems, whether in electrical circuits, the brain, or life itself, evolve from simple components. He emphasized the importance of breaking problems down to their basics before building up. Through his guidance, I gained a deeper appreciation for physics, coding, and the art of problem-solving. His mentorship has profoundly shaped my approach to science.

A well-deserved thanks goes to my friends Lindsey Salay, Janis Hesse, and Walter Gonzalez for their friendship throughout this journey, a friendship given freely, without asking anything in return. To the entire Tsao lab, current and past members, Joseph Wekselblatt, Francisco Luongo, Nicole Schweers, Varun Wadia, Krithika Mohan, Nate Dolensek, Dasheng Bi, Ruichang Sun, Audo Flores, Pinglei Bao, Liang She, Tomo Sato, Lu Liu, Erin Koch, Jialiang Lu, Yuelin Shi, Irene Caprara, Hongsun Guo, Shi Chen, Seoyoung Ahn, and David Chung, I am deeply grateful for their unwavering support and the valuable exchange of ideas that has enriched my lab life.

Finally, I want to thank Caltech and its scientific community. As a little kid, I dreamed of attending this place of science nestled among palm and orange trees, and it is a rare thing when dreams come true.

ABSTRACT

In the mammalian visual system, photons captured by the retina are transformed into meaningful internal percepts of surroundings through a hierarchy of interconnected visual areas. Understanding the representation of visual information at each node of the hierarchy has been a central quest of visual systems neuroscience over the past 50 years. The primate visual system, with its over two dozen distinct areas broadly organized into a dorsal stream for visuo-motor transformations and a ventral stream for object recognition, has served as the gold standard for studying the organization of the visual system. Recent advances in artificial neural networks modeled on the primate visual system for object recognition have prompted the question, *is hierarchical representation necessary, and if so, can we observe it across all highly visual mammalian species?* Hierarchical organization appears to be a key architectural principle of both artificial and biological networks, enabling stepwise construction of a structured and compact representation from raw sensory input. Here we present a series of efforts to determine the cortical organization and connectivity of the tree shrew visual system and directly compare to that of the primate. This cross-species study sheds light on the evolution and mechanisms of vision in a close relative of primates. Using high-density Neuropixels recordings, we demonstrate that the tree shrew ventral visual pathway exhibits primate-like hierarchical processing, with progressively larger receptive fields, increasing response latencies, and enhanced selectivity for complex stimuli along the visual pathway. Area V2 in the tree shrew performs key functions similar to those of the primate inferotemporal (IT) cortex. Specifically, V2 contains strongly face-selective cells, supports a complete representation of high-level object space, and achieves the most accurate object identity decoding and reconstruction among all tree shrew visual areas. Yet we also found significant differences from the canonical template for hierarchical organization observed in the primate, including maintenance of relatively small, focal receptive fields throughout the hierarchy, and better decoding of latent variables in late deep neural network (DNN) layers by area V2 compared to other areas.

The hierarchical organization of the visual system describes the arrangement of areas but does not reveal how information flows between them. Understanding the type of processing carried out at each node raised the next question of whether information that is transmitted across nodes is differentiated between feedforward and feedback connections. To explore this, we combined electrical microstimulation and extracellular recordings to identify the directionality of projections which is applicable in various species. We used this technique to first study the connections between the first two nodes of the tree shrew cortical hierarchy, V1 and V2. We found that V2 feedback neurons carry a full visual representation on par to other V2 cells. These feedback neurons were distinct with regards to their spatial features, including distinct locations and sizes of their receptive fields. We also found that both feedforward and feedback V2 neurons were modulated by perceptual conflict arising when distinct textures were presented to each eye, suggesting they could refine V1 processing to perceptual inconsistencies.

These studies provide insights into how the tree shrew visual system generates object representations through a hierarchy of interconnected nodes, employing strategies adapted to its cortical constraints. In addition, by combining electrical microstimulation with electrophysiology we set the foundation for cross-species studies to determine the role of feedforward and feedback processing along the visual hierarchy. Together, this work reveals conserved principles of visual processing across species while showcasing unique adaptations in the tree shrew, offering insights into the evolutionary origins and functional organization of the primate visual system.

Table of Contents

Chapter I

INTRODUCTION

Motivation

Long before written language, humans sought to make sense of their surroundings by transferring what they saw into lines etched on cave walls. Drawing is one of humanity's most primal acts as a way to engage with and understand the world through vision, a means of distilling the complexities of the visual world into its simplest, most meaningful forms. This act of simplification resonates with a fundamental aspect of how the brain processes visual information. Just as an artist reduces the infinite variations of light and shadow into a few strokes of charcoal, the brain abstracts and interprets sensory input into something comprehensible. Vision, however, is not passive reception but an active process, a continuous negotiation between what the eye perceives and what the mind internally constructs.

How does the visual system construct an object percept? This ability underpins how we interact with the world, guiding essential functions like navigation and social behaviors. Both biological and artificial neural networks demonstrate the capacity for object recognition[1,2], potentially achieving it through diverse mechanisms or converging on similar, efficient solutions. Studying how evolution has shaped the visual abilities of different organisms, and comparing them with manmade artificial architectures, can offer a unique perspective to uncover common core principles for object vision.

Principles of hierarchical organization for form vision

In mammals, the cortical visual system is hierarchically organized, composed of a series of interconnected areas that transforms simple pixels into detectable objects[3,4]. This cortical functional architecture has been most extensively studied in the primate visual system. Beginning with the retina, visual information passes through successive stages of processing in the brain, where increasingly abstract features are extracted. The primate visual system exemplifies this hierarchy, with distinct areas such as V1, V2, and the inferotemporal (IT) cortex specializing in different aspects of perception, from the detection of simple edges to the recognition of complex objects and faces[5–8]. This deeply layered structure has long been regarded as a pinnacle of evolutionary adaptation, enabling the high visual acuity and object recognition that primates rely on for survival.

Understanding what gives rise to visual perception requires uncovering the progressive visual transformation at each stage in the hierarchy. The receptive field is a key concept central to visual processing, defined as the spatial region where a stimulus triggers a neuron's response[9]. An individual neuron's receptive field properties are shaped by the type, number, and arrangement of its inputs. The receptive field conveys information about visual space (i.e. the location within the visual field) as well as a particular feature within that space (e.g. motion, color, shape), with increasing complexity at each stage of the hierarchy. Early in the primate visual pathway, such as in primary visual cortex (V1), neurons have small, precise receptive fields, enabling them to detect localized, fine details like edges and orientations. As information flows through the visual hierarchy, neurons with small receptive fields tuned to simple features are combined to form neurons with larger receptive fields that encode more complex and abstract features[10,11]. At later stages of the primate visual hierarchy, receptive fields become very large and spatially-invariant to support critical visual functions, such as the ability to recognize objects regardless of variations in size, position, or orientation[12]. This hierarchical building of representations is the cornerstone for the primate visual system which is highly adapted for sophisticated object recognition.

At the pinnacle of the primate ventral visual stream lies an area, the inferotemporal cortex (IT), that constitutes the apex of specialized object processing[13]. Within IT, topographically organized regions contain cells specialized in identifying specific object categories. The most well-known example is the face patch system, a series of interconnected regions that contains almost entirely face-selective cells[14]. Collectively, the activity of these neurons forms a generative model of faces, enabling reconstruction of an individual face with remarkable precision using signals from just a small number of neurons[15]. Importantly, this encoding strategy extends beyond face recognition and generalizes to the broader IT cortex, where similar computational principles apply to other object categories. As information progresses beyond IT, the level of abstraction increases, enabling representations that encode object familiarity and begin to blur the line between object perception and object memory.

The concept of hierarchy has profoundly shaped our understanding of the architecture of the mammalian visual system for over 50 years, serving as a foundational principle for the development of advanced multi-layered computational networks[3,4]. Artificial neural networks (ANNs), particularly deep learning (DL) models, draw inspiration from this biological architecture, mimicking the hierarchical and modular organization of the brain[2,16]. These artificial systems are designed to emulate the stepwise abstraction observed in the brain, with early layers detecting basic features such as edges or textures and deeper layers capturing complex patterns and relationships. This principle, rooted in biological vision, has driven remarkable advancements in machine vision and pattern recognition, enabling tasks such as object detection, face recognition, and even artistic image synthesis. While DL models serve as powerful tools for understanding perception and cognition, they also act as computational frameworks for modeling brain activity. However, the relationship between artificial systems and biological brains raises important questions. Do these models accurately reflect the underlying neural mechanisms, or are their representations simply convergent solutions to shared computational problems? That is, do similarities between artificial and biological representations reveal insights about the brain itself, or do they primarily teach us about the multidimensional space that both systems are designed to navigate?

Probing the flow of visual information along the hierarchy

The unique neural computations performed by neurons are largely shaped by their inputs that convey information about the outside world. Previous advances in neuroscience techniques have established a foundation for recording neural activity from large population of neurons in the visual system[17–19]. However, they lack the important ability of establishing directionality of neural signals. This is important since the physiological response properties of visual neurons are not only inherited from feedforward inputs but are also shaped by their feedback connections. Numerically, there are as many feedback connections as feedforward along the many stages of the hierarchy[20–22]. The function of these neurons based on their projection targets has predominately eluded most studies as it is difficult to identity the directionality of surveyed neurons using electrophysiological recordings.

Traditionally, experimentally challenging methods were employed to identify projection neurons, including leveraging the use of electrical microstimulation to unambiguously determine the directionality of the axons[23]. However, these techniques were limited by the fact that antidromic identification of neurons that satisfies the standard criterion of passing a collision test (i.e. transmission failure due to action potential collision[19]) produces very low yields[24,25]. Thus, these techniques were not often used, leaving a potential gap in knowledge about the function of neurons in the visual cortex based on their reciprocal connectivity. However, with the recent

advent of high-density silicon probes, such as Neuropixels, this now offers a significant advantage by allowing for simultaneous recordings of many hundreds of neurons which can be combined with electrical microstimulation to identify large numbers of projection cells.

Electrical microstimulation is a well-established technique that is useful for probing the functional circuitry that offers to advantage of broad applicability across species. This combined technique has been able to measure neuronal interactions in real time among rodents, birds and non-human primates. By affecting neural activity via the voltage gradient that neurons maintain across their membranes; a current passed outside of cells can change this voltage and trigger neuronal responses[26,27]. Electrical microstimulation offers improved biological compatibility as opposed to photostimulation, because it does not require artificially expressing proteins on the neural membrane (i.e., channelrhodopsin[28]) to excite the cell. Rather, electrical stimulation exploits the existing electrical properties of neurons and does not require the introduction of foreign proteins, a process that greatly increases the experimental time and may produce unintended biological effects. In addition to speeding up the experimental set-up, electrical microstimulation is not constrained by areas of viral expression, so the site of electrical microstimulation can be moved easily between brain areas during the same experiment. Finally, recent advances in genetic and viral tools allow for mapping and monitoring neurons based on their projection targets, such as phototagging and calcium imaging[29]. While this has been widely available in rodent models, it is much less feasible in other species such as tree shrews and primates in which the tools are still in the early phases of implementation[30].

Advantages of electrical microstimulation notwithstanding, some challenges that need to be overcome include electrical artifacts that often overwhelm the recording setup during electrical microstimulation, causing a spatiotemporal “blur” of the neural activity[31,32]. Microstimulation artifacts can occur because of capacitive crosstalk between electrodes or because of large electric fields from the stimulation site reaching the recording site. Because the extracellular voltage signals associated with action potentials are several orders of magnitude weaker than the pulses required for successful stimulation, these fields can easily overwhelm the recording amplifiers and result in long-lasting oscillations in the recording circuitry, which make it impossible to record neuronal activity for tens of milliseconds after the stimulus. To overcome this, some recording designs have implemented analog switches that are disabled during stimulation to prevent these large voltages from reaching the amplifier. However, the current generation of silicon probes does not allow for this method, thus promoting the need to design optimal solutions.

Overcoming these technical challenges and employing these techniques in species such as tree shrews and primates will open possibilities to understand and dissect at a circuit level the contributions of different projection neurons to signal processing along the visual hierarchy. Notably, applications include elucidating the functional role of feedback connections in visual processing, which has continued to remain a fundamental mystery in vision science. In the visual cortex, there have been many suggested roles for top-down feedback connections, including regulating spatial properties (i.e. receptive fields) or providing specialized feature-specific signals. Prior work has suggested that these connections could either be suppressing or facilitating depending on the location of inputs and the cell types that they synapse onto. Anatomical tracing experiments have shown excitatory feedback connections can project to either excitatory or inhibitory neurons[33]. Also, recent findings in the mouse visual cortex has shown that based on whether the source of the feedback is aligned or offset from the target in visual space, it will result in excitation or suppression in V1 respectively[34,35]. This raises additional questions such as ‘what types of visual information is sent back to earlier areas and how do they contribute to visual perception?’. One hypothesis is that they function to conflict or ambiguity in visual stimuli[36–38]. Feedback connections from higher areas are also thought to enhance visual processing by

modulating and contextualizing responses at earlier stages, such as due to global changes in arousal and attention. Overall, this bidirectional flow of information offers additional challenges to understanding how vision works by greatly increasing the complexity of computations performed at each node since they must be considered in the framework of the complexities of a recurrent dynamical system.

An evolutionary perspective: using new animal models for vision

The act of seeing is shaped by the evolutionary imperative to navigate, recognize, and interact with the environment. A functional visual system must detect and parse light, discern edges and movement, and interpret shapes and forms. These tasks require neural circuits finely tuned to the demands of the organism's environment, reflecting a balance between computational efficiency and ecological necessity. But this raises an intriguing question: do all mammals with complex visual behavior rely on deep hierarchies? Or can evolution arrive at alternative solutions?

Historically, most vision studies have been performed in humans, macaque monkeys, and cats. In the last 15 years, there has been a growing interest in using rodents as a model organism to parse the mechanisms underlying visual processing [39,40]. While mice have relatively poor visual acuity 10 times lower than primates, they offer significant advantages in terms of applications for labeling and manipulating defined cell types. This has prompted a new wave of interest in identifying a model species that can combine the genetic tractability of mice with the visual acuity that more closely resembles the primate. For those reasons, tree shrews have become increasingly used in visual neuroscience to advance towards the goal of understanding how vision works.

Tree shrews are highly visual, diurnal mammals. Whether they should be classified as primates, insectivores, or, as has eventually become established, *scandentia*, has garnered some debate. Like most mammals, they are dichromats[41]. Not only are they diurnal, but their visual systems are highly adapted to diurnal life including an overrepresentation of cones comprised of ~95% of the total photoreceptors[41]. In humans, cones mediate high-spatial acuity and thus also likely contribute to the high visual acuity of tree shrews. The anatomy and physiology of the tree shrew visual system has therefore attracted much attention. Anatomical tracing studies have suggested the existence of a greatly expanded visual cortex with regions thought to be akin to many corresponding regions in the primate. However, there has been an overall lack in studies to systematically determine whether these regions are only ostensible analogs due to matching anatomical structure or actual functional homologs and whether they conform to the primate-like hierarchical organization. If so, this would place tree shrews as an ideal species for studying the evolution of form vision.

Evolutionarily positioned between rodents and primates, comparative studies of the tree shrew visual systems stand to provide new insights into core concepts for complex visual processing. Investigating homologies and evolutionary relationship between the visual system of diverse animals offers to teach us fundamental principles of biological systems and how they are built. Vision has evolved independently across countless lineages, from the compound eyes of insects to the pinhole eyes of mollusks, and the camera-like eyes of mammals. The eye itself provides insights into the evolutionary change across species that selects for characteristics optimal for various ethological niches. A clear example lies in the regional differences in retinal ganglion cell distributions typically corresponding to where in visual space high detail vision is needed, such as the differential specializations in aerial versus ground dwelling organisms. Certain computational challenges remain universal. Many building blocks (i.e. photoreceptors, retinal

ganglion cells) and neural computations (i.e. motion, color, orientation) are largely preserved as they support similar functions to guide perception and interactions with the outside world.

Outline

This thesis explores how the brains of different visual species that are widely different in size and have adapted for different environments can take in photons and reconstruct a stable visual percept of an object essential for survival. By comparing across different neural networks, can we uncover core principles that are conserved across evolution and across species that emerge from simple interactions among interconnected areas?

In Chapter II, we explore this question by performing a comparative analysis of visual systems across mammals and artificial neural networks. Using electrophysiological recording with Neuropixels probes, we examined the visual responses in many tree shrew brain areas to a large stimulus set. We showed that cells in area V2 exhibit selectivity for complex objects that is generally not found in area V2 of the primate. This study revealed that animals with fewer visual brain areas have processing networks that are compressed and potentially exhibit more recurrency than feed-forward processing.

Chapter III presents the technical advancements we achieved to combine the use of Neuropixels recordings with electrical stimulation to identify neurons based on their projections. By overcoming challenges associated with electrical artifacts, we applied this technique to probe the role of feedforward and feedback processing in tree shrews' primary visual cortex (V1) and secondary visual cortex (V2). We identified distinct classes of V2 neurons based on the directionality of their projections. We observed that 1) feedback neurons in V2 convey similar visual information compared to other V2 neurons. 2) Feedback exhibits distinct features such as smaller receptive fields and spatial offsets relative to V1 inputs. 3) Finally, both feedforward and feedback neurons encode information about perceptual conflict, such as when distinct images are presented to each eye. These results highlight the role of feedback in refining V2 processing to resolve perceptual inconsistencies.

Finally in Chapter IV we introduce future direction about how feedforward and feedback connections in the compressed hierarchy in the tree shrew compares to the multi-stage processing in primates to achieve object recognition. We offer insights into this by investigating this question in the primate face patch system that contains extensive feedforward and feedback connections. Using electrical microstimulation and Neuropixels recordings, we investigated the role of these connections by stimulating two face patches at opposite ends of the hierarchy. Our observations point to a role in feedback in potentially refining representations under uncertain conditions. Future work exploring cross-species comparisons for the distinct role of feedforward and feedback in form vision offers to advance our understanding of visual perception.
space.

References

1. Bao, P., She, L., McGill, M. & Tsao, D. Y. A map of object space in primate inferotemporal cortex. Nature 583, 103–108 (2020).
2. Yamins, D. L. K. et al. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. U. S. A. 111, 8619–8624 (2014).
3. Van Essen, D. C. & Maunsell, J. H. R. Hierarchical organization and functional streams in the visual cortex. Trends Neurosci. 6, 370–375 (1983).
4. Felleman, D. J. & Van Essen, D. C. Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cereb. Cortex (1991).
5. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. J. Physiol. 195, 215–243 (1968).
6. Gross, C. G. & Schonen, S. D. Representation of visual stimuli in inferior temporal cortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. 335, 3–10 (1992).
7. Mishkin, M., Lewis, M. E. & Ungerleider, L. G. Equivalence of parieto-preoccipital subareas for visuospatial ability in monkeys. Behav. Brain Res. 6, 41–55 (1982).
8. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311 (1997).
9. Sherrington, C. The integrative action of the nervous system. J. Nerv. Ment. Dis. 34, 801–802 (1907).
10. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160, 106–154 (1962).
11. Wallisch, P. & Movshon, J. A. Structure and function come unglued in the visual cortex. Neuron vol. 60 195–197 (2008).
12. Kobatake, E. & Tanaka, K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J. Neurophysiol. 71, 856–867 (1994).

13. Hesse, J. K. & Tsao, D. Y. The macaque face patch system: a turtle's underbelly for the brain. Nat. Rev. Neurosci. 21, 695–716 (2020).
14. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H. & Livingstone, M. S. A cortical region consisting entirely of face-selective cells. Science 311, 670–674 (2006).
15. Chang, L. & Tsao, D. Y. The Code for Facial Identity in the Primate Brain. Cell 169, 1013-1028.e14 (2017).
16. Zador, A. et al. Catalyzing next-generation Artificial Intelligence through NeuroAI. Nat. Commun. 14, 1597 (2023).
17. Trautmann, E. M. et al. Large-scale high-density brain-wide neural recording in nonhuman primates. bioRxiv (2023) doi:10.1101/2023.02.01.526664.
18. Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. Nature 551, 232–236 (2017).
19. Bishop, P. O., Burke, W. & Davis, R. Single-unit recording from antidromically activated optic radiation neurones. J. Physiol. 162, 432–450 (1962).
20. Briggs, F. Role of Feedback Connections in Central Visual Processing. Annual Review of Vision Science 6, 1–22 (2020).
21. Gilbert, C. D. & Sigman, M. Brain states: top-down influences in sensory processing. Neuron 54, 677–696 (2007).
22. Wang, Q. & Burkhalter, A. Area map of mouse visual cortex. J. Comp. Neurol. 502, 339–357 (2007).
23. Moore, T. & Armstrong, K. M. Selective gating of visual signals by microstimulation of frontal cortex. Nature 421, 370–373 (2003).
24. Movshon, J. A. & Newsome, W. T. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J. Neurosci. 16, 7733–7741 (1996).

25. El-Shamayleh, Y., Kumbhani, R. D., Dhruv, N. T. & Movshon, J. A. Visual response properties of V1 neurons projecting to V2 in macaque. Journal of Neuroscience 33, 16594–16605 (2013).
26. Histed, M. H., Bonin, V. & Reid, R. C. Direct Activation of Sparse, Distributed Populations of Cortical Neurons by Electrical Microstimulation. Neuron 63, 508–522 (2009).
27. Histed, M. H. & Miller, E. K. Microstimulation of frontal cortex can reorder a remembered spatial sequence. PLoS Biol. 4, 826–835 (2006).
28. Deisseroth, K. Optogenetics. Nat. Methods 8, 26–29 (2011).
29. Luo, L., Callaway, E. M. & Svoboda, K. Genetic dissection of neural circuits: A decade of progress. Neuron 98, 865 (2018).
30. Tremblay, S. et al. An open resource for non-human primate optogenetics. Neuron 108, 1075-1090.e6 (2020).
31. Semework, M. Microstimulation: Principles, techniques, and approaches to somatosensory neuroprosthesis. Crit. Rev. Biomed. Eng. 43, 61–95 (2015).
32. Butovas, S. & Schwarz, C. Spatiotemporal Effects of Microstimulation in Rat Neocortex: A Parametric Study Using Multielectrode Recordings. J. Neurophysiol. 90, 3024–3039 (2003).
33. Siu, C., Balsor, J., Merlin, S., Federer, F. & Angelucci, A. A direct interareal feedback-to-feedforward circuit in primate visual cortex. Nat. Commun. 12, 4911 (2021).
34. Fişek, M. et al. Cortico-cortical feedback engages active dendrites in visual cortex. Nature 617, 769–776 (2023).
35. Keller, A. J., Roth, M. M. & Scanziani, M. Feedback generates a second receptive field in neurons of the visual cortex. Nature 582, 545–549 (2020).
36. Ekstrom, L. B., Roelfsema, P. R., Arsenault, J. T., Bonmassar, G. & Vanduffel, W. Bottom-up dependent gating of frontal signals in early visual cortex. Science 321, 414–417 (2008).
37. Friston, K. The free-energy principle: a rough guide to the brain? Trends Cogn. Sci. 13, 293–301 (2009).

38. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2, 79–87 (1999).
39. Huberman, A. D. & Niell, C. M. What can mice tell us about how vision works? Trends Neurosci. 34, 464–473 (2011).
40. de Vries, S. E. J. et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. Nat. Neurosci. 23, 138–151 (2020).
41. Jacobs, G. H. & Neitz, J. Spectral mechanisms and color vision in the tree shrew (Tupaia belangeri). Vision Res. 26, 291–298 (1986).

Chapter II

A compressed hierarchy for visual form processing in the tree shrew

Authors: Frank F. Lanfranchi[1,2*], Joseph Wekselblatt[2], Daniel A. Wagenaar[2,3], Doris Y. Tsao[1,4*†]

Affiliations:
[1]Dept of Molecular and Cell Biology, UC Berkeley
[2]Division of Biology and Biological Engineering & Computation and Neural Systems, Caltech
[3]Caltech Neurotechnology Laboratory
[4]Howard Hughes Medical Institute

*Corresponding author. Email: flanfran@caltech.edu, dortsao@berkeley.edu
†Lead Contact

Keywords:
tree shrew; functional hierarchy; receptive field; ventral stream; visual cortex; Neuropixels; deep networks; face cells.

Abstract:
Our knowledge of brain processes governing vision is largely derived from studying primates, whose hierarchically stratified visual system[1] inspired the architecture of deep neural networks[2]. This raises questions about the evolutionary origins and universality of such hierarchical structures. Here, we examined the large-scale functional organization for vision in one of the closest living relatives to primates, the tree shrew. We performed Neuropixels recordings[3,4] across multiple cortical and thalamic areas spanning the tree shrew ventral visual system while presenting a large battery of visual stimuli in awake tree shrews. We found an increase in receptive field size, response latency, and selectivity for naturalistic textures compared to spectrally-matched noise[5] moving anteriorly along the tree shrew visual pathway, consistent with a primate-like hierarchical organization[6,7]. Surprisingly however, we found that tree shrew area V2 already harbored a high-level representation of complex objects. First, V2 carried a complete representation of a high-level object space[8]. Second, V2 activity supported the most accurate object reconstruction among all tree shrew visual areas. Finally, starting in area V2, we found strongly face-selective cells akin to those reported in primate inferotemporal cortex[9] which could support face identity decoding. Overall, the results shed new light on the origins of the primate visual system, showing how core computational principles of visual form processing are conserved, yet hierarchically compressed, by evolution in a small but highly visual mammal.

INTRODUCTION

The ability to recognize objects is fundamental to the survival of visual animals. The primate ventral stream has provided a cornerstone for studying how objects are processed in the brain[10,11]. A hallmark of the primate ventral stream is hierarchical organization[12], which is strikingly mirrored by deep neural networks (DNNs) trained on object recognition [8,13]. This raises the question: *Is hierarchical representation necessary, and if so, can we observe it across all highly visual mammalian species?* Do the brains of cats, tree shrews, diurnal bats, and other highly visual mammals all possess object recognition pathways following the same hierarchical architecture and functional properties as the primate brain? Investigating visual processing across different mammalian species promises to provide a deeper understanding of general principles for object vision expressed across evolution.

Over a decade ago, the mouse visual system began to attract strong interest, driven by the wealth of tools available for mouse neural circuit dissection[14,15]. However, the mouse's low visual acuity and limited cortical territory dedicated to vision[16] make it a non-ideal organism for studying hierarchical brain mechanisms underlying object recognition. The tree shrew has attracted growing interest as a model to study visual processing[17] due to its high visual acuity (>10x that of rodents)[18], greatly expanded visual cortex[19], and excellent ability to perform visually-guided behavioral tasks compared to the mouse[20,21]. The tree shrew visual system is differentiated into at least nine distinct anatomical visual cortical areas [19]. Tree shrews also possess a high degree of functional specialization within primary visual area (V1), with an orderly arrangement of orientation selective columns[22,23] and a prominent second visual area (V2), albeit with a large-scale topographic organization that differs from that of the primate[24]. Lesion studies suggest a rough correspondence between tree shrew extrastriate areas anterior to V2 and primate IT cortex: ablations of large portions of extrastriate cortex in the temporal lobe of the tree shrew produce deficits in pattern discrimination and object vision similar to the effect of IT lesions in primates[19,25,26]. However, to our knowledge there have been no electrophysiological studies of functional properties of extrastriate visual areas beyond V2 in the tree shrew.

Here, we aim to identify the cortical organization and coding principles underlying visual object representation across the entire tree shrew ventral stream. We performed large-scale electrophysiological recordings using multiple Neuropixels probes along five tree shrew ventral visual areas as well as the pulvinar. We confirmed hallmarks of hierarchical organization found in primates including increased receptive field size and response latency[27] as well as increased selectivity for naturalistic textures compared to spectrally matched noise[5] moving anteriorly along the tree shrew visual pathway. However, unexpectedly, we found that tree shrew area V2 already harbored a highly sophisticated representation of complex objects, comparable to that of macaque IT cortex. V2 cells carried a full representation of a high-level object space and axis tuning to features of this space. Furthermore, tree shrew area V2 supported the best stimulus reconstruction among all tree shrew visual areas, matching the performance obtained by pooling activity across all visual areas. Remarkably, among cells selective for different sectors of a deep network–derived object space, we found strongly face-selective cells that could support identity decoding. These cells were found throughout the tree shrew visual pathway starting in area V2. This is remarkable because tree shrews are not social animals, living in isolated monogamous pairs[28]; the finding of face cells in this species challenges the prevailing view that face cells evolved under pressures of social communication [29,30]. Overall, the results suggest a compressed, multi-stage hierarchy in the tree shrew in which representations previously observed in the primate are realized at a much earlier stage of visual processing.

RESULTS

We selected a set of areas that span the tree shrew ventral stream to target for the investigation of hierarchical visual processing (**Fig. 1a**). We chose primary (V1) and secondary (V2) visual areas as architectonically distinct regions that participate in early visual processing and are mutually connected[31,32]. We further selected the temporal posterior area (TP) as an intermediate step along the ventral visual processing stream. This region, located immediately anterior to V2, receives dense projections from V1, V2, and temporal anterior (TA) and temporal dorsal (TD) areas which have visuo-motor functions. Importantly, unlike TA and TD, TP does not project to primary motor cortex, and thus is considered a more purely visual area[19]. Finally, at the anterior end of the tree shrew ventral stream, we surveyed three subregions that may be homologous to macaque inferior temporal cortex based on anatomical location: temporal-inferior (TI), temporal intermediate (ITi), and inferotemporal rostral (ITr) areas. Lesions to TI and ITi cause drastic impairments in visual form detection[25]. ITr is known to receive inputs from both visual and auditory

cortex[19], but its visual functional properties have never been explored. Due to difficulty in distinguishing the border between TI and ITi, we grouped results for these two regions and refer to them as "TI-ITi." Since many temporal areas receive direct thalamic input[33], we also recorded in the dorsal visual portion of the pulvinar (Pulv). We performed retrograde tracing experiments to guide target location for electrophysiological recording (**Extended Data Fig. 1a, b**).

To characterize the visual responses of neurons across V1, V2, TP, TI-ITi, ITr, and Pulv, we performed electrophysiological recordings using Neuropixels probes in awake tree shrews (**Fig 1b**). In all experiments, animals were head-fixed and restrained in front of a monitor while we presented a battery of visual stimuli, including local sparse noise, static gratings, naturalistic textures and noise, and images of faces and other objects. At the end of experiments, probe locations were marked with DiI and targeting was confirmed with histology (**Fig. 1c**). We labeled a cell as visually responsive if it responded to any of the classes of visual stimuli we tested (*Methods*). We found many well-isolated single units in each area (**Fig. 1d**), with some inter-area differences in the fractions of cells that responded to visual stimuli (ANOVA, $F_{5,17} = 4.0$, $p < 0.014$; **Fig. 1e**). In particular, relatively fewer TI-ITi cells were visually responsive compared to V2 cells.

We first mapped the receptive fields of neurons along the tree shrew ventral pathway using a locally sparse noise stimulus (*Methods*). We measured receptive fields of single neurons by fitting a gaussian distribution to the two-dimensional matrix of spike counts at each location of the visual field; ON and OFF receptive fields were computed separately using responses to white and black squares, respectively. Cells with ON and/or OFF receptive fields were clearly present in all areas except TP (**Fig. 1f**). Surprisingly, this included the two most anterior areas TI-ITi and ITr; in contrast, corresponding areas in the anterior primate temporal lobe show largely spatially invariant responses[34,35].

Within individual recordings, receptive field positions were concentrated in a small portion of the screen, corresponding to our electrode penetration of the cortical surface. **Fig. 1g** shows receptive fields of all cells recorded in one example session for each area. Surprisingly, clustering of receptive fields was apparent in all areas studied, suggesting that even TI-ITi and ITr, though located at the anterior end of the tree shrew ventral stream, maintain retinotopic organization.

To assess the hierarchical relationships between the recorded areas, we first examined two classic metrics of hierarchical level: receptive field size and the latency of visually-evoked spikes. We found that the receptive field size systematically increased from posterior to anterior (**Fig. 1h**). We also calculated the half-peak latencies for each unit in each area and found that latencies increased from V1 to V2 to ITr (**Fig. 1i**, *Methods*). The hierarchy predicted by the increase in receptive field sizes was broadly consistent with the hierarchy predicted by the increase in latencies (**Fig. 1j**).

In the primate visual cortex, early visual areas are strongly tuned to low-level features such as orientation and spatial frequency, while later areas are tuned to more complex object features[7,36–38]. We therefore examined tuning to orientation and spatial frequency across the tree shrew ventral visual pathway using static gratings (**Fig. 2a**). We found that the proportion of visually-responsive neurons (cf. **Fig. 1e**) that specifically responded to gratings was the highest in V1 and V2 (~55% and ~65% respectively) and lowest in TI-ITi (**Fig. 2b**). Tuning to orientation, spatial frequency, and spatial phase of example cells from V2 and ITr illustrates the different types of tuning we observed to these variables across tree shrew visual areas (**Fig. 2c**). Overall, orientation tuning was most prevalent in V1 and V2 (Tukey analysis after ANOVA, $F_{5,1099} = 25.9$, $p < 10^{-24}$, **Fig. 2d**), while spatial frequency tuning was also prevalent in ITr (Tukey analysis ANOVA, $F_{5,1099} = 19.4$, $p < 10^{-17}$, **Fig. 2e**). These findings are roughly consistent with those found

in the primate ventral stream, where orientation tuning is especially prominent in early visual areas[39].

So far, V2 responses appeared largely similar to V1 responses, raising the question of whether V2 performs any unique function distinct from that of V1. Previous studies in the macaque have identified sensitivity to higher-order statistical dependencies in naturalistic textures as a unique signature of area V2[5]. We therefore asked if tree shrew extrastriate areas show a similar specialization for naturalistic texture. To address this, we recorded from all six areas while presenting naturalistic textures and spectrally matched synthetic noise images (**Fig. 2f**, *Methods*). Among all areas recorded, V2 had the largest proportion of cells responding to the texture and/or noise stimuli (**Fig. 2g)**. Examination of the time course of population responses revealed the strongest difference between responses to naturalistic textures versus spectrally matched noise in area V2, followed by V1, ITr, and TI-ITi, while no or only slight differences were found in the remaining areas (**Fig. 2h**). In V2, the difference persisted for the duration of the stimulus. Interestingly, although responses in V1 commenced well before those in V2 (cf. **Fig. 1i**), the modulation between texture versus noise occurred later in V1 (at 90 ms) than in V2 (at 45 ms), suggesting that the modulation in V1 may be a result of feedback from V2. This idea is further supported by the observation that activity in V2 encoded texture family identity earlier than that in V1 (**Fig. 2i**).

A central purpose of the visual hierarchy is to recognize and categorize objects to guide vital behaviors like navigation, foraging, or mating. To understand the mechanisms for high-level object representation in the tree shrew ventral stream, we presented a rich stimulus set consisting of 1593 images of animals, body parts, faces, and everyday objects (*Methods*); this stimulus set was previously used to characterize tuning in macaque IT cortex, enabling a direct comparison to object recognition mechanisms in primates[8]. Stimuli were adjusted to match the receptive field location of neurons recorded (*Methods*). As we show below, responses to these complex object stimuli suggest a very different picture from that so far, where the tree shrew appears to harbor a primate-like hierarchy with increasing receptive field size and latency (**Fig. 1**) and decreasing prominence of orientation tuning (**Fig. 2**) moving anteriorly. Instead, the key conclusion from the next set of results derived from the complex object stimuli is the unique role of area V2 in representing high-level object features.

Response rasters of example cells vividly showcase the diversity in object selectivity throughout the tree shrew ventral stream (**Fig. 3a**). Among the six areas we recorded from, a similar proportion of visually-responsive cells responded to the object stimuli across V2, TP, TI-ITi, and Pulv (**Fig. 3b**), with a much larger proportion of visually-responsive cells in TI-ITi responding to object stimuli compared to gratings (cf. **Fig. 2b**). This increased responsiveness to complex objects compared to simple gratings is consistent with temporal areas being situated at a higher stage in the tree shrew ventral pathway. For each cell, we quantified the “explainable variance” as the part of the variance in response that could be attributed to the identity of the visual stimuli presented (rather than to stochastic trial-to-trial variation) (*Methods*). After V2, the explainable variance in responses to these complex object stimuli decreased dramatically (**Fig. 3c**). Overall, these results suggest that even though anterior tree shrew areas were relatively more responsive to complex objects than to gratings, their responses to these complex object stimuli were not highly reproducible from trial-to-trial. When controlling for low-level features, we did not observe a large fraction of the variance explained was due to these features including luminance, contrast, and spatial frequency (**Fig 3c** and **Extended Data Fig. 2**).

To better understand the nature of the neural code used by each area, we modeled neural responses using AlexNet[40], an 8-layered DNN trained on object recognition (**Fig. 3d**). Single IT

neurons in the macaque monkey have been found to be optimally explained by an axis model, wherein each cell linearly projects incoming stimuli onto a preferred axis in an DNN-derived feature space[8,13]. In primates, the preferred axes of the cell population span a relatively low-dimensional basis for the feature space (e.g., in face patches, just 50 dimensions can yield veridical reconstructions[41]). To test whether this principle also holds in the tree shrew, we calculated the preferred axis of each neuron in the six areas, using the first 50 principal components (PCs) from AlexNet layer FC6. We focused on FC6 in order to delve into the question of whether tree shrew cortex represents a high-level object space, as in macaque IT cortex[8]. We found that visual cells in all areas of the tree shrew ventral stream showed ramp-shaped tuning along their preferred axes (**Fig. 3e,** *Methods*). Moreover, cells showed flat tuning along their principal orthogonal axis (i.e., longest axis orthogonal to the preferred axis; **Fig. 3f**, *Methods*).

Previous studies in primates revealed that early layers of AlexNet and other DNNs explain neuronal activity in early retinotopic visual areas better, whereas late DNN layers explain inferotemporal responses better[8,13]. We investigated whether a similar pattern holds for the tree shrew ventral stream. We regressed firing rates of single cells to 50 feature dimensions of a given AlexNet layer (*Methods*) and asked which layer of AlexNet best explained the variance in neuronal responses. For one representative cell in V2, AlexNet layer Conv4 best explained its responses (**Fig. 4a**). Across the V2 population, the middle layers of AlexNet (Conv4, Conv5) had more explanatory power for cells' responses than either the early or late layers (**Fig. 4b**).

To compare explanatory power of different AlexNet layers across brain areas, we calculated the sum across cells within each area of the variance explained by the various AlexNet layers, and normalized these sums by the sum across cells of their explainable variance (*Methods*). This analysis revealed that early visual areas V1 and V2 were best explained by early layers Conv3 to Conv5, whereas TI-ITi and ITr were best explained by FC6 (**Fig. 4c**). However, the absolute explanatory power of AlexNet was lower for the higher cortical areas (**Extended Data Fig. 3a, b**), consistent with the lower explained variance by image identity in anterior areas (cf. **Fig. 3c**). This could suggest that AlexNet may not be expressive enough to capture the response properties of tree shrew IT as they have been suggested to be multimodal[19]

To explore which feature axes encoded the most variance in neural responses in each area, we plotted how much variance was explained by individual feature PCs from AlexNet layer FC6. We found that in general, earlier PCs explained more variance in neural responses, with some variability across areas (**Fig. 4d**). Conversely, we also analyzed how well specific features of AlexNet FC6 could be decoded from population activity in different tree shrew visual areas (**Fig. 4e**). Again, we found that early PCs were most strongly represented in the tree shrew visual hierarchy, with substantially better decoding of FC6 features from V2 activity than any other area. This is consistent with the fact that FC6 explained more of the variance of V2 activity than of other areas (**Extended Data Fig. 3c**). Thus, even though V2 was best explained byConv4 and Conv5 features, while TI-ITi and ITr were best explained by FC6 features, nevertheless FC6 features were better represented in V2 than in these more anterior areas. Additionally, we measured whether low-level features correlate with FC6 PCs and found that they were very small (**Extended Data Fig. 3d, e**).

Given the high performance for decoding AlexNet FC6 features using V2 activity, we asked whether activity in V2 might be sufficient to reconstruct objects using small neural populations, as has previously been shown in monkey IT cortex[8]. To this end, we took a large auxiliary dataset comprising 15901 images and passed them through AlexNet. Whichever image yielded activations in FC6 closest to the reconstructed activation from a given area of the tree shrew brain was considered the reconstructed image (**Extended Data Fig. 3f**). We performed reconstructions using 100 randomly selected cells from each area, to control for cell number. Consistent with our

results on parameter decoding (**Fig. 4e**), which were optimal in V2, we found that images reconstructed from V2 closely resembled the original images, while images reconstructed from V1 or TI-ITi were less close (**Fig. 4f**). To quantitatively compare reconstruction accuracy across areas, we calculated the distance between FC6 activations reconstructed based on neural responses to each image in each area and the actual FC6 response to that image, normalized to the theoretical best decoding distance (*Methods*). We found that these normalized decoding distances were significantly smaller, on average, in V2 than in other areas, and matched the performance obtained using all neurons pooled across all areas (Tukey analysis after ANOVA, $F_{5,9552}$ = 235.2; **Fig. 4g**).

The primate IT cortex is organized into subregions containing cells that respond maximally to images from specific categories, e.g., faces. Moreover, these cells are clustered into discrete anatomical regions by image category [42–44]. Such category-selective regions can be explained by a normative account in which IT cortex represents a general "object space" defined as the span of the first two PCs of the AlexNet FC6 representation of the 1593 object images [8,45]. Specific sectors in this space correspond to distinct categories of objects, such as images of faces, fruits, and animals (**Fig. 5a**).

Does the tree shrew visual cortex, like primate IT cortex, contain regions specialized for representing distinct sector(s) of object space? To address this question, we projected the preferred axes of all recorded cells down onto the same 2D object space (**Fig. 5b**). We found that preferred axes of cells in area V2 spanned all four quadrants, whereas preferred axes of cells in other areas largely fell into quadrants I and III. Given the localization of different categories to different sectors of this object space, one prediction is that individual tree shrew cells should be selective for distinct categories. This prediction was confirmed by analysis of response rasters of individual cells. Cells with preferred axes in the face quadrant were indeed strongly face selective (**Fig. 5c**). Some face cells also responded to other round shapes, while others showed strong selectivity only for faces. We also found cells selective for spiky, elongated objects (quadrant I), round inanimate objects (quadrant II), and spiky animate objects (quadrant IV; **Fig. 5d** and **Extended Data Fig. 4a, b**).

We next sought to compare how each area at the population level responds to the images corresponding the different quadrants of object space. **Extended Data Fig. 4c** shows the projection of all 1593 images onto the 2D object space, with the 100 most preferred images for each area indicated by colored dots. The majority of the top-100 preferred images of TI-ITi fell in the quadrant that contains faces (quadrant III), while none of the top-100 preferred images of V1 and V2 fell in this quadrant. To address whether cells selective for different categories are spatially clustered within individual tree shrew visual areas, we replotted the preferred axes of cells using color coding to represent each cell's depth along the Neuropixels probe (**Extended Data Fig. 4d**). This revealed no obvious topographical organization within any area for visual preference.

Faces—particularly human faces, which all our face stimuli were composed of—are not known to have special behavioral importance to tree shrews[28]. Thus, we performed additional analyses to confirm that the cells really were face selective. We introduced a quantitative face selectivity index, defined as the difference between responses to faces and all other objects, for each individual cell (*Methods*). This confirmed small populations of highly face-selective cells (t-score ≥ 15) in most areas starting in area V2, with the highest percentage in TI-ITi and pulvinar (**Fig. 5e**).

The primate IT cortex is specialized for object recognition and has remained the cornerstone for studying form processing. To facilitate direct comparisons with our tree shrew dataset, we

performed large-scale recordings using NHP Neuropixels probes in macaque monkeys. We presented the same 1593 object stimuli while recording from V2, posterior IT (ITpost), and anterior IT (ITant) from two monkeys per area (**Fig. 6a-c**). We found the explainable variance in responses to complex object stimuli increased from primate V2 to ITant (**Fig. 6d**), whereas they peaked in tree shrew V2 (c.f. **Fig. 3c**). Image reconstruction performance also increased along the primate hierarchy (c.f. **Fig. 4g**), whereas it was most optimal in tree shrew V2 (**Fig. 6e**). In contrast to tree shrews (c.f. **Fig. 5e**), we did not observe strongly face-selective cells in primate V2 (**Fig. 6f**). As expected, the number of face cells in primate ITpost and ITant were much higher. Of note in one of the recordings from ITpost, the probe was partially targeted one of the defined face patches resulting in a higher proportion of face cells.

Lastly, we tested whether the primate and tree shrew cell populations could decode individual face identity or object identity. To this end, we trained a classifier to decode the individual identity either of 200 faces, or of 200 general objects, using the neural activity of randomly sampled subpopulations of cells from a given area (**Fig. 6g,** *Methods*). In tree shrews, all areas yielded above-chance decoding for both faces and objects except area TP. When we selectively used only face cells, decoding performance of face identity increased dramatically, especially in temporal areas. Most strikingly, decoding performance in tree shrew area V2 was much better than in all other areas for both face and object decoding. Strikingly, we found the decoding performance of object identity from the primate V2 was far lower than that from the tree shrew V2 (**Fig. 6f**). In fact, tree shrew V2 was similar to that of primate posterior IT. As expected, the primate anterior IT that sits at the apex of the primate ventral visual stream outperformed all other areas.

A hallmark of the primate ventral stream is increasing invariance moving anteriorly along the temporal lobe[8,34,46]. We used responses to the 1593 objects across different tree shrew visual areas to ask whether invariance increases along the tree shrew ventral pathway as well (even though this image set was not designed to test view invariance directly). Here, we leveraged the power of deep network models trained on large image sets to predict responses of visual neurons[47,48]. We first confirmed that *in the macaque*, a deep network-based model of V2 and IT cells trained on neuronal responses to the 1593 images effectively predicted responses of the same cells to a second image set consisting of 51 objects presented at 24 different views (**Extended Data Fig. 5a, b, c**;). We then used these predicted responses to objects at different views to compute a predicted invariance index for each macaque cell (*Methods*). This analysis revealed a strong correlation between predicted and actual invariance indices, with the predicted overestimating and corresponding to the upper bound of the actual (**Extended Data Fig. 5e**). Thus, this confirmed that responses to the 1593 image set were sufficient to capture view invariance of cells in the macaque. We then applied this approach to compute predicted invariance indices across different tree shrew areas. In the tree shrew, the distributions of invariance indices from different areas were highly overlapping (**Extended Data Fig. 5f**) Thus, unlike in the macaque, increasing view invariance was not observed for predicted invariance in the tree shrew. As has been previously shown in macaque (Doris XXX) and rats (XXX), direct testing within each area is needed to determine whether view invariance is a hallmark of the tree shrew visual pathway (**Fig. 6h)**.

DISCUSSION

Hierarchical processing is a central principle of object representation in artificial neural networks and in the primate visual system. Here, we sought to determine the extent to which the ventral visual pathway of the tree shrew, a highly visual mammal that is one of the closest existing relatives to the primate[49], is also organized hierarchically. To this end, we performed large-scale simultaneous electrophysiological recordings spanning six nodes of the tree shrew visual pathway

starting from early visual cortex and culminating in the presumed homolog to primate inferotemporal cortex (**Fig. 1**). In support of hierarchical organization, we found that higher-level areas in the tree shrew ventral visual stream showed an increase in receptive field size, response latency, selectivity for naturalistic textures compared to spectrally matched noise, and proportions of single cells selective for faces.

However, what was most striking were the deviations of tree shrew visual system organization from the canonical template for hierarchical organization observed in the primate. While receptive field sizes were the largest in TI-ITi and ITr, they were comparatively small (mean = 5.6° and 6.1° respectively) relative to the many large spatially-invariant receptive fields of primate anterior IT cells (spanning on average 10-20°[34])). Further, decoding of latent variables of late AlexNet layers (which in the primate most closely match IT) was best using activity from tree shrew area V2 (**Fig. 5f**). Indeed, by decoding FC6 features from area V2, we could obtain reasonable reconstructions of objects using only small cell populations (**Fig. 4f**). Area V2 harbored a full representation of a high-level object space previously shown to be represented by primate IT cortex[8] (**Fig. 5b**). In particular, area V2 contained a population of highly face-selective cells that represented the face quadrant of this object space and supported face identity decoding better than any other area (**Fig. 5f**). Finally, DNN-predicted indices of view invariance were as high in tree shrew area V2 as in more anterior areas (**Extended Data Fig. 4**), with the caveat that this needs to be confirmed with direct measurement of view invariance in future studies. Overall, these results suggest a substantially shallower visual hierarchy in the tree shrew compared to the primate, with tree shrew area V2 performing many of the functions of primate IT cortex. An open question for future research is: to what extent does tree shrew area V2 encompass all of the functions of primate IT cortex? For example, in object recognition tasks, can task performance be entirely explained by activity of V2 cells?

As a direct comparison, we performed the same experiments across three homologous regions in the macaque, including V2, IT anterior and IT posterior. Strikingly, this comparison further supports our finding that the tree shrew contains a compressed hierarchy that performs many of the functions akin to primate IT in tree shrew V2, and importantly, primate V2 does not show these characteristics. Our stimulus set was originally tailored for primate object recognition and, therefore, facilitated direct comparisons to primates. However, future work is needed to expand these findings with additional stimulus sets such as those that include ethologically relevant objects, view invariance, binocular disparity and multimodal stimuli. In fact, studies in rats using highly controlled visual morphed objects designed to match luminosity across transformations led to the discovery of properties of object processing in the rat visual area LL and TO including view invariance. Such comparisons in the tree shrew would further shed light into evolutionarily conserved properties for object vision that places this species functionally between the primates and rodents.

Our findings challenge the current focus on modeling mechanisms for high-level vision almost exclusively with deep networks[50,51]. Computationally, deep networks aid in the challenge of sequentially disentangling image features that are important for discrimination from orthogonal features such as orientation and size[46,52]. However, it is possible that V2 itself harbors a deep network implemented through local circuits, e.g., via a recurrent network that, when temporally unrolled, can be identified with a multi-layer feedforward network but would require less neurons to implement[53]. Future work may investigate this possibility by analyzing local dynamics of feature selectivity within V2.

We did not find any striking, qualitative difference in the complexity of visual processing between area V2 and more anterior areas in the tree ventral stream (TP, TI-ITi, and ITr). This raises the

question, what is functionally distinct about these more anterior areas? One possibility is that these areas are involved in multi-sensory integration, consistent with the existence of anatomical connection with the pulvinar and other higher order auditory cortical areas[19]. Thus, these more anterior areas may inherit their visual tuning from V2 without extensive further processing, and their primary function may be to integrate this visual tuning with tuning for other sensory features.

The finding of face cells in the tree shrew was particularly surprising. The existence of face cells in primates has long been thought to be related to the importance of faces for primate social communication[54]. However, recent evidence argues that such specializations may arise from more fundamental principles concerning how IT cortex represents a general object space[8,45,55,56]. Facial communication is not known to be of ethological importance to tree shrews (which live in isolated monogamous pairs, with social recognition mediated by olfactory cues[28]). Thus the finding of face cells in tree shrew visual cortex supports the view that such cells can readily emerge from encoding of general dimensions of image variation, even in the absence of evolutionary pressures related to face-based social communication.

The tree shrew offers exciting advantages as a model organism for studying high-level vision given its tractability for genetic and viral-mediated circuit approaches and its highly developed visual system. In particular, the tree shrew visual system appears to be more sophisticated than that of the mouse, evinced by preferential responses to naturalistic textures, face-selective responses, and the existence of five distinct visual cortical areas (V1, V2, TP, TI-ITi, ITr) with increasing receptive field size and latency. Our study provides a new roadmap for exploring visual circuits in this non-traditional species and illuminates how evolution adapted brains of different sizes for effective representation of the visual world.

CONTRIBUTIONS

Conceptualization, FFL, DAW, DYT; software, DAW; formal analysis, FFL, DAW; experiments, FFL; tree shrew colony, FFL, JW; surgeries, FFL, JW; resources, DYT; writing—original draft, FFL, DYT; writing—review and editing, all authors.

ACKNOWLEDGEMENTS

This work was supported by NIH Pioneer Award Grant DP1-NS083063, the Howard Hughes Medical Institute, and the Tianqiao and Chrissy Chen Institute for Neuroscience at Caltech. We thank David Fitzpatrick for help establishing a tree shrew colony and for technical assistance; Nicole Schweers for technical support; F. J. Luongo for advice and assistance with electrophysiology and the members of the Tsao laboratory; George Mountoufaris, Walter Gonzalez, and Lindsey Salay for critical comments as well as feedback on the manuscript.

METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experimental procedures were approved by the Caltech Institutional Animal Care and Use Committee and conformed to local and US National Institutes of Health guidelines, including the US National Institutes of Health Guide for Care and Use of Laboratory Animals. Tree shrews (*Tupaia Belangeri*) used in this study (n=5), both male and female, were 6 months to 2.5 years old and weighted between 150 to 300 g. Animals were singly housed in a 12-hour light/dark cycle in the animal room. Their food and water aliquots were given *ad libitum*.

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

EXPERIMENT

Surgeries

Tree shrews were injected with a preoperative dose of dexamethasone (5 mg/kg, subcutaneously (s.c.)) and mannitol (1 mg/kg, s.c.) to reduce swelling. Animals were anesthetized with a cocktail of fentanyl, midazolam, and dexdomitor (FMD, fentanyl 0.05 mg/kg, midazolam 5.0 mg/kg, dexdomitor 0.25 mg/kg, s.c.), shaved and positioned into a stereotaxic frame. Topical lidocaine gel (2%) was applied on the head and ears to prevent discomfort from ear-bars and eye lubricant was used to maintain hydration and clarity of eyes during surgical procedures. Levels of anesthesia, breathing, SpO2, and heart rate were monitored throughout the entire procedure and body temperature was maintained with a heating pad at 37.5° C. An incision on the scalp was performed and both skin and muscles were retracted. The exposed skull was levelled using the stereotaxic device with respect to bregma and lambda (pitch, roll, and yaw). After alignment, locations of the craniotomies for electrophysiological recordings were marked on the skull and a custom stainless steel headplate was secured to the skull using clear C&B Metabond (Parkell). A layer of Kwik-Cast (World Precision Instruments) was added on top the skull and a 3D printed custom cap was secured to the headplate to protect the brain and keep debris out. The anesthesia was reversed with an injection of atipamezole-flumazenil (atipamezole 1.25 mg/kg, flumazenil 0.25 mg/kg, s.c.) and the animal was recovered for at least 3 days before following procedures and recordings. One day before electrophysiological recordings in a new brain location, tree shrews were once again anesthetized and monitored as described above. Using the marked locations on the skull, small (up to 1.5mm of diameter) craniotomies were drilled and durotomy was performed. Through a small hole situated anterior of bregma, a 32 AWG chlorinated silver wire (A-M system) with a pre-soldered gold pin was implanted just above the brain surface and cemented to the skull to provide chronic grounding. A drop of silicone oil (30,000 cSt, Aldrich) was

added over the holes to prevent the brain from drying, a new layer of Kwik-Cast was applied on top of it, and the 3D printed custom cap secured to the headplate. Anesthesia was reversed and the animal was recovered as previously described.

Electrophysiological recordings

All electrophysiological recordings were made using high channel-count, silicon, “Neuropixels 1.0” probes configured to always acquire from the first 384 electrodes closest to the tip, providing a 3.84 mm of tissue coverage. The reference and the ground contacts on the Neuropixels probes were permanently soldered together. Recordings were made using an external reference configuration achieved by connecting the probe reference to the chronically implanted silver wire on the skull; conductivity was routinely checked before recording with a multimeter. Each Neuropixels was mounted on a 3-axis micromanipulator (New Scale Technologies) that was in turn mounted on the underside of a semicircular platform, allowing simultaneous insertion of up to four probes at different angles. Before the first insertion of a probe in a new location, DiI (1 mM in ethanol) was used to coat the shank, allowing subsequent probe track localization during ex vivo imaging. Neural signals were acquired at 30 kHz using Open Ephys software [57]. After the tip of each probe touched the surface of the brain, they were lowered to target at an average speed of 100 µm/min to avoid damage and let them settle for 15 minutes after reaching the target depth. Cameras were used to monitor animals during experiments and ensure a continuative viewing of the visual stimuli presented during neural signals acquisition. After each recording experiment, probes were slowly retracted and immersed in 1% Tergazyme solution to remove tissue and silicone oil residues.

Injections

To trace the inputs to TP and ITr, intracranial injections were performed as described in the surgical procedure as above. The retrograde tracer cholera toxin subunit β was injected into TP (CTβ-488) and into ITr (CTβ-594) using a pulled glass capillary (World Precision Instruments) and a pressure injector (Micro4 controller, World Precision Instruments), at a flow rate of 50 nl/ min. The tracer was delivered at two depths, -- mm below the cortical surface, to ensure adequate spreading. Stereotaxic injection coordinates were based on the Zhou and Ni Tree Shrew brain atlas[58] (TP, anterior–posterior: −6.43 mm, medial–lateral: ±8 mm, dorsal–ventral: −5.5 mm; ITr, anterior–posterior: −1.54 mm, medial–lateral: ±8 mm, dorsal–ventral: −5 mm relative to Bregma). Perfusions and histology were performed seven days following injections.

Histology

After electrophysiological recordings or tracer expression, histological verification was performed for all tree shrews. Tree shrews were given Ketamine + Xylazine and perfused transcardially with 0.9% saline, followed by 4% paraformaldehyde (PFA) in 1X PBS. Brains were extracted and post-fixed overnight in 4% PFA at 4°C. The brains were then transferred to 30% sucrose for cryoprotection and sectioned coronally at 100 µm on a cryostat (Leica Biosystems). Sections were washed with 1xPBS and then incubated for 30 minutes at room temperature in in DAPI/PBS (0.5 µg/ml) for counterstaining. Sections were then mounted on slides and imaged with an epifluorescence microscope (Olympus VS120).

VISUAL STIMULATION

Visual stimuli presentation

Visual stimuli were generated and presented using custom Python scripts. Head-fixed tree shrews passively viewed a battery of visual stimuli displayed using a ViewSonic monitor (70x39 cm, 60 Hz refresh rate, 1,920 × 1,080 pixels). The monitor was centered in front of the animals at 25 cm distance. Stimuli were presented at 3 Hz, 167 ms of image presentation interleaved with 167 ms

of a grey screen. Three classes of visual stimuli were used in each experiment: static gratings, naturalistic textures and noise, and 1593 objects. In addition, "local sparse noise" stimuli were used to map neurons' receptive fields.

Local Sparse Noise

The screen was divided into a grid of 4x3 squares. In consecutive frames (100 ms), sparse white or black dots (5 degrees square) were presented, one dot in each grid square. The locations of the dots within each rectangle were pseudo-randomly distributed to avoid spurious correlation between distant parts of the visual field[27]. To avoid interference between reconstruction of "On" and "Off" RFs, each presented stimulus frame comprised either all black or all white dots on a grey field. A reduced version of this stimulus (with fewer frames) was used at the beginning of each experiment and analyzed immediately to allow placement of "faces and objects" stimuli in the centroid of the receptive fields for that recording session.

Before the main recording session, a shorter block of local sparse noise stimuli was presented to the animal and results were analyzed to find the approximate centroid of the receptive field of recordable neurons, to serve as the location for subsequent presentation of faces and objects stimuli.

Static Gratings

We presented full field sinusoidal gratings, varying in orientation (6 evenly spread angles), spatial frequency (5 values between 0.1 and 1.6 cycles/degree), and phase (4 positions), for a total of 120 different stimulus conditions. Each image was presented 5 times.

Naturalistic textures and noise

We presented images from two subclasses: naturalistic textures and a control set comprising spectrally matched noise. The naturalistic texture images were organized as 15 families of 5 similar images. Texture images reproduced the statistical dependencies found in natural texture scenes[1]. Each of the 150 images in the stimulus set was presented 5 times. We used two types of visual stimuli similar to ones previously used in primate studies: one set consisted of 15 families of texture images, each comprising 5 closely related image samples of the same texture. These images reproduced statistical dependencies found in natural texture scenes[5,59]. A control set consisted of noise images spectrally matched to each of the texture families.

Faces and objects

We presented images from two subclasses: 1,392 objects and animals from www.freepngs.com, and 201 faces from the FEI database[8], for a total of 1,593 images. Each image was presented 10 times. Images were presented at the previously determined center of the receptive field of recordable cells and sized to cover 20 degrees of the visual field, which covered the majority of the recorded neurons' receptive fields.

DATA ANALYSIS

Preprocessing and Spike sorting

Neural signals from electrophysiological recordings were preprocessed by subtracting the median calculated within each group of 24 channels from the data to eliminate common-mode noise. The median subtracted data was sent to Kilosort2 which in addition to the group median subtraction applied a high-pass filter (150-Hz), followed by whitening in blocks of 32 channels. All spike data were sorted using the off-line spike sorting algorithm Kilosort2[60]. The cluster automatically labelled by Kilosort algorithm as "good" was in turn manually curated by hand and further analyzed with Phy2.

Visually responsive cells

A cell was deemed responsive to a particular class of stimuli (either gratings, textures, and noise, or faces and objects) if its average firing rate in the 100 ms following stimuli of that class exceeded the expectation value based on a Poisson model trained on the firing rate in the 50 ms before all the stimuli of that class. To be included in the "responsive fraction" in figures 3b, 4b, and 5b, a cell's average response had to exceed the baseline by at least 5 standard deviations. For the "faces and objects," the total time elapsed between the first and the last of the 10 blocks of visual presentations was so long that stability of responses was a concern. Accordingly, we additionally preprocessed these data to analyze only those blocks in which the responses were stable for a given cell. For each block, we extracted the average waveform of all the spikes from the given cell and calculated its peak-to-peak amplitude. We then picked the third largest amplitude among the blocks and set an amplitude threshold at 0.6x this value. We counted for each block the number of individual spikes with amplitudes exceeding this threshold. We calculated the mean and standard deviation of these counts among blocks, and excluded from analysis any block in which the count was over two standard deviations below the mean. In all cases except figure 1e, results are expressed as a percentage of visually responsive cells, i.e., of cells that respond to any of the stimulus classes.

Receptive field analysis

The receptive field size, amplitude, and quality was obtained by first calculating a 2D histogram of spike counts at each of 576 locations on the monitor (32 x 18 matrix). We modeled these histograms as a 2D Gaussian peak on top of a constant baseline. To prevent overfitting, the shape of the Gaussian was forced to be circular rather than elliptic. A cell was considered to possess an ("ON" or "OFF") receptive field if the number of spikes within the Gaussian peak exceeded expectation from a null model. Specifically, we calculated the expected number of spikes that would be elicited by ("ON" or "OFF") stimuli within a 10-degree radius from the center of the Gaussian under the null model of the baseline as well as the actual number of spikes elicited by stimuli within that same area. The number of standard deviations by which the actual number of spikes exceeded the null expectation was considered the "quality" of the RF. Only cells with RF quality greater than 5 were considered to possess an RF.

Half-peak latency

For each neuron, we calculated the average response to all the gratings and texture/noise stimuli as a function of latency after stimulus onset. We found the peak value in this peristimulus time histogram (PSTH) and kept only cells in which the peak exceeded the 99.75%th percentile of the Poisson distribution predicted from baseline firing. The "half-peak latency" of a cell was defined as the latency at which its response first exceeded a threshold set halfway between its baseline firing rate and the peak.

Preferred orientation and spatial frequency

We analyzed responses to gratings in terms of orientation and spatial frequency of the gratings. First, we grouped trials by orientation and fitted a modified Von Mises distribution to the response data for each neuron, where the orientation space of 0° to 180° was treated as the full period for the purpose of the distribution. The preferred orientation of a cell was the centroid of the fitted distribution. Separately, we grouped trials by spatial frequency. We fitted a Gaussian distribution to the responses in log-frequency space. The preferred spatial frequency of a cell was the center of the fitted distribution. In both figures 3e and f, only cells were included in the count where the amplitude of the (Von Mises or Gaussian) peak was at least 0.5 times the average firing rate during all gratings responses of the given cell.

Percentage variance explained

In Figs. 3g, 3h, 4d, 4i, and 5c, we plot the percentage of variance in neuronal activity that is explained by various discrete or categorical variables. (We treat orientation and spatial frequency as discrete variables here.) First, we calculate the total variance (V_{total}) in neuronal activity across trials for a given neuron. Then we regress the activity onto the categorical variable (k – 1 additional degrees of freedom, where k is the number of values the variable can attain) and calculate the residual variance ($V_{residual}$). By definition, the explained variance is the difference between total and residual variance, and the plotted percentage is: $100\% \times (V_{total} - V_{residual}) / V_{total}$. In fig. 4d, 4i, and 5c, we separately add up the total and residual variances for all the neurons in an area before normalizing. (That is, we plot the percentage of all the variance in the area that is explained by the variable, rather than the average across cells of the explained variance for each cell.)

Explainable variance

To derive an upper bound on the maximum fraction of variance that could theoretically be explained by the DNN, we calculated the "explainable variance" of the neuronal responses as the split-half reliability of those responses using the Spearman-Brown formula $2\rho / (1+ \rho)$ applied to the correlation between the responses to the same image in one half of the trials to the other half[8].

Preferred axis (AlexNet)

We extracted the activations of the 4096 units in layer FC6 of AlexNet in response to each of the 1593 images and performed principal component analysis to reduce the 4096-dimensional space down to 50 dimensions. For each cell, we calculated which axis in this space captured the largest fraction of the variance in its responses to all but 10% of the images. We then calculated the projection onto the found axis of the remaining images. We repeated this process 10 times, each time keeping a different set of images as a test set. This yielded projection values for every image in the data set. The average of the 10 axes found is the cell's overall "preferred axis." We defined bins over the projection values and calculated the average response of the cell to all the images in that bin. Each pixel in the matrices in Fig. 5e represents one such average.

Principal orthogonal axis (AlexNet)

As a control, we took the first principal component of the AlexNet responses and, for each cell, projected it down to the hyperplane orthogonal to that cell's preferred axis. This we call the "principal orthogonal axis" for that cell.

Decoding AlexNet activation from neural activity

We repeatedly selected 100 cells at random from a given area, calculated PCs from their responses to image stimuli, and used those to regress the individual PCs of AlexNet. After repeated sampling, we calculated the average fraction of the AlexNet PC's variance that was explained by the neural data from a given area.

Object reconstruction and normalized decoding distance

Image reconstructions were performed as previously described[8,61]. To generate images that reflect the features encoded in the neural responses, we passed into AlexNet images from an auxiliary database comprising a much larger set of 15901 images, none of which was previously shown to the animal. For each stimulus image presented to the animal, the feature vector decoded from the neural activity was compared to the feature vectors of the larger auxiliary stimulus set. We defined the "reconstructed image" as the image in the auxiliary dataset with the smallest Euclidean distance to the decoded feature vector of the original image.

Given that the auxiliary images used for reconstruction did not include any of the objects shown to the animals (limiting how good the reconstruction can be), we computed a 'normalized decoding

distance' to quantify the reconstruction accuracy for each object. We first used the Moore–Penrose pseudoinverse to transform the predicted features from neuronal data back into the space of AlexNet layer FC6 activations. Next, we calculated the Euclidean distance between these pseudoinverted predicted features and the actual AlexNet FC6 activations deriving from the presented images. We normalized this distance by the theoretical best decoding distance, i.e. the distance between the actual AlexNet FC6 activation and the back projection of the 50D PCA output of AlexNet FC6 (again using the Moore–Penrose pseudoinverse). Thus, the normalized decoding distance for an image is:

$$Normalized\ decoding\ distance = \frac{\left|V_{recon} - V_{original}\right|}{\left|V_{bestpossiblerecon} - V_{original}\right|}$$

where $\mathbf{v}_{recon}$ is the feature vector reconstructed from neuronal responses, $\mathbf{v}_{original}$ is the feature vector of the image presented to the animal, and $\mathbf{v}_{best\ possible\ recon}$ is the feature vector of the best possible reconstruction. A normalized distance of one means that the reconstruction has found the best solution possible.

Face selectivity

For every cell we quantified its selectivity to faces by calculating the t-score between its responses to faces, and its responses to the rest of the images. In all areas an unexpectedly large number of cells had t-scores far beyond the expected null distribution.

In Extended Data Fig 4b, we sorted cells in both macaque IT targets by face selectivity index calculated as FSI = (mean $response_{face}$ - mean $response_{non\ face\ objects}$) / (mean $response_{face}$ + mean $response_{non\ face\ objects}$).

Face and object identity decoding

We trained Gaussian Naïve Bayes classifiers to extract object or face identities from sets of 200 images based on the activity of variously sized subsets of neurons from a given area, using 9 out of 10 repeated presentations of each image for training, and the remaining presentations for testing. We calculated the significance of the decoding accuracy over chance by Monte Carlo resampling both over repeated presentations of the same image, and over random subsamplings of neurons from the population of each given area. We calculated the average performance of each sample of neurons across different selections of training sets.

Invariance index

The Invariance Index was calculated as the mean of the Pearson correlation coefficients between the frontal view and 23 other non-frontal views, averaged across identities. This index reflects the consistency of the neuronal response to the same stimulus presented under different conditions.

REFERENCES

1. Felleman, D. J. & Van Essen, D. C. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb. Cortex* (1991).
2. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
3. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
4. Trautmann, E. M. *et al.* Large-scale high-density brain-wide neural recording in nonhuman primates. *bioRxiv* (2023) doi:10.1101/2023.02.01.526664.
5. Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.* **16**, 974–981 (2013).
6. Kobatake, E. & Tanaka, K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* **71**, 856–867 (1994).
7. Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**, 215–243 (1968).
8. Bao, P., She, L., McGill, M. & Tsao, D. Y. A map of object space in primate inferotemporal cortex. *Nature* **583**, 103–108 (2020).
9. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H. & Livingstone, M. S. A cortical region consisting entirely of face-selective cells. *Science* **311**, 670–674 (2006).
10. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
11. Logothetis, N. K. Visual Object Recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
12. Tanaka, K. Inferotemporal Cortex and Object Vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
13. Yamins, D. L. K. *et al.* Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8619–8624 (2014).

14. Seabrook, T. A., Burbridge, T. J., Crair, M. C. & Huberman, A. D. Architecture, Function, and Assembly of the Mouse Visual System. *Annu. Rev. Neurosci.* **40**, 499–538 (2017).
15. de Vries, S. E. J. *et al.* A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat. Neurosci.* **23**, 138–151 (2020).
16. Huberman, A. D. & Niell, C. M. What can mice tell us about how vision works? *Trends Neurosci.* **34**, 464–473 (2011).
17. Savier, E., Sedigh-Sarvestani, M., Wimmer, R. & Fitzpatrick, D. A bright future for the tree shrew in neuroscience research: Summary from the inaugural Tree Shrew Users Meeting. *Zool Res* **42**, 478–481 (2021).
18. Petry, H. M., Fox, R. & Casagrande, V. A. Spatial contrast sensitivity of the tree shrew. *Vision Res.* **24**, 1037–1042 (1984).
19. Wong, P. & Kaas, J. H. Architectonic subdivisions of neocortex in the tree shrew (Tupaia belangeri). *Anat. Rec.* **292**, 994–1027 (2009).
20. Schumacher, J. W., McCann, M. K., Maximov, K. J. & Fitzpatrick, D. Selective enhancement of neural coding in V1 underlies fine-discrimination learning in tree shrew. *Curr. Biol.* **32**, 3245-3260.e5 (2022).
21. Li, C., McHaney, K. M., Sederberg, P. B. & Cang, J. Tree Shrews as an Animal Model for Studying Perceptual Decision-Making Reveal a Critical Role of Stimulus-Independent Processes in Guiding Behavior. *eNeuro* **9**, (2022).
22. Bosking, W. H., Zhang, Y., Schofield, B. & Fitzpatrick, D. Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex. *Journal of Neuroscience* 2112–2127 (1997).
23. Lee, K. S., Huang, X. & Fitzpatrick, D. Topology of on and off inputs in visual cortex enables an invariant columnar architecture. *Nature* **533**, 90–94 (2016).
24. Sedigh-Sarvestani, M. *et al.* A sinusoidal transformation of the visual field is the basis for periodic maps in area V2. *Neuron* **109**, 4068-4079.e6 (2021).

25. Killackey, H., Snyder, M. & Diamond, I. T. Function of striate and temporal cortex in the tree shrew. *J. Comp. Physiol. Psychol.* **74**, Suppl 2:1-29 (1971).
26. Mishkin, M., Ungerleider, L. G. & Macko, K. A. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* **6**, 414–417 (1983).
27. Siegle, J. H. *et al.* Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92 (2021).
28. Emmons, L. *Tupai: A Field Study of Bornean Treeshrews*. (University of California Press, 2000).
29. Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R. & Livingstone, M. S. Seeing faces is necessary for face-domain formation. *Nat. Neurosci.* **20**, 1404–1412 (2017).
30. Perrett, D. I., Rolls, E. T. & Caan, W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp. Brain Res.* **47**, 329–342 (1982).
31. Sesma, M., Casagrande, V., Kaas, J. H. & Vac, A. Cortical Connections of Area 17 in Tree Shrews. *J. Comp. Neurol.* **230**, 337–351 (1984).
32. Lyon, D. C., Jain, N. & Kaas, J. H. *Cortical Connections of Striate and Extrastriate Visual Areas in Tree Shrews*. vol. 401 109–128 (1998).
33. Petry, H. M. & Bickford, M. E. The Second Visual System of The Tree Shrew. *Journal of Comparative Neurology* vol. 527 679–693 Preprint at https://doi.org/10.1002/cne.24413 (2019).
34. Ito, M., Tamura, H., Fujita, I. & Tanaka, K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* **73**, 218–226 (1995).
35. Freiwald, W. A. & Tsao, D. Y. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330**, 845–851 (2010).
36. Tolhurst, D. J. & Movshon, J. A. Spatial and temporal contrast sensitivity of striate cortical neurones. *Nature* **257**, 674–675 (1975).

37. Gross, C. G., Bender, D. B. & Rocha-Miranda, C. E. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* **166**, 1303–1306 (1969).
38. Tanaka, K. Neuronal mechanisms of object recognition. *Science* **262**, 685–688 (1993).
39. David, S. V., Hayden, B. Y. & Gallant, J. L. Spectral receptive field properties explain shape selectivity in area V4. *J. Neurophysiol.* **96**, 3492–3505 (2006).
40. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017).
41. Chang, L. & Tsao, D. Y. The Code for Facial Identity in the Primate Brain. *Cell* **169**, 1013-1028.e14 (2017).
42. Hesse, J. K. & Tsao, D. Y. The macaque face patch system: a turtle's underbelly for the brain. *Nat. Rev. Neurosci.* **21**, 695–716 (2020).
43. Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. Visual properties of neurons in inferotemporal cortex of the Macaque. *J. Neurophysiol.* **35**, 96–111 (1972).
44. Perrett, D. I., Hietanen, J. K., Oram, M. W. & Benson, P. J. Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **335**, 23–30 (1992).
45. Doshi, F. R. & Konkle, T. Cortical topographic motifs emerge in a self-organized map of object space. *Sci Adv* **9**, eade8187 (2023).
46. DiCarlo, J. J. & Cox, D. D. Untangling invariant object recognition. *Trends Cogn. Sci.* **11**, 333–341 (2007).
47. Ponce, C. R. *et al.* Evolving Images for Visual Neurons Using a Deep Generative Network Reveals Coding Principles and Neuronal Preferences. *Cell* **177**, 999-1009.e10 (2019).
48. Bashivan, P., Kar, K. & DiCarlo, J. J. Neural population control via deep image synthesis. *Science* **364**, (2019).
49. Fan, Y. *et al.* Genome of the Chinese tree shrew. *Nat. Commun.* **4**, (2013).

50. Suzuki, M., Pennartz, C. M. A. & Aru, J. How deep is the brain? The shallow brain hypothesis. *Nat. Rev. Neurosci.* **24**, 778–791 (2023).
51. Nayebi, A. *et al.* Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation. *PLoS Comput. Biol.* **19**, e1011506 (2023).
52. Ma, Y., Tsao, D. & Shum, H.-Y. On the Principles of Parsimony and Self-Consistency for the Emergence of Intelligence. *arXiv [cs.AI]* (2022).
53. Sutskever, I. Training recurrent neural networks. (University of Toronto, CAN, 2013).
54. Barraclough, N. E. & Perrett, D. I. From single cells to social perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 1739–1752 (2011).
55. Vinken, K., Prince, J. S., Konkle, T. & Livingstone, M. S. The neural code for “face cells” is not face-specific. *Sci Adv* **9**, eadg1736 (2023).
56. Shi, Y. *et al.* Rapid, concerted switching of the neural code in inferotemporal cortex. *bioRxiv* (2023) doi:10.1101/2023.12.06.570341.
57. Siegle, J. H. *et al.* Open Ephys: An open-source, plugin-based platform for multichannel electrophysiology. *J. Neural Eng.* **14**, (2017).
58. Zhou, J.-N. & Ni, R.-J. *The Tree Shrew (Tupaia Belangeri Chinensis) Brain in Stereotaxic Coordinates*. (Springer, 2016).
59. Portilla, J. & Simoncelli, E. P. *A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients*. vol. 40 49–71 (2000).
60. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris, K. Fast and accurate spike sorting of high-channel count probes with KiloSort. *Adv. Neural Inf. Process. Syst.* **29**, 4448–4456 (2016).
61. Wadia, V. S. *et al.* A shared code for perceiving and imagining objects in human ventral temporal cortex. *bioRxiv* (2024) doi:10.1101/2024.10.05.616828.

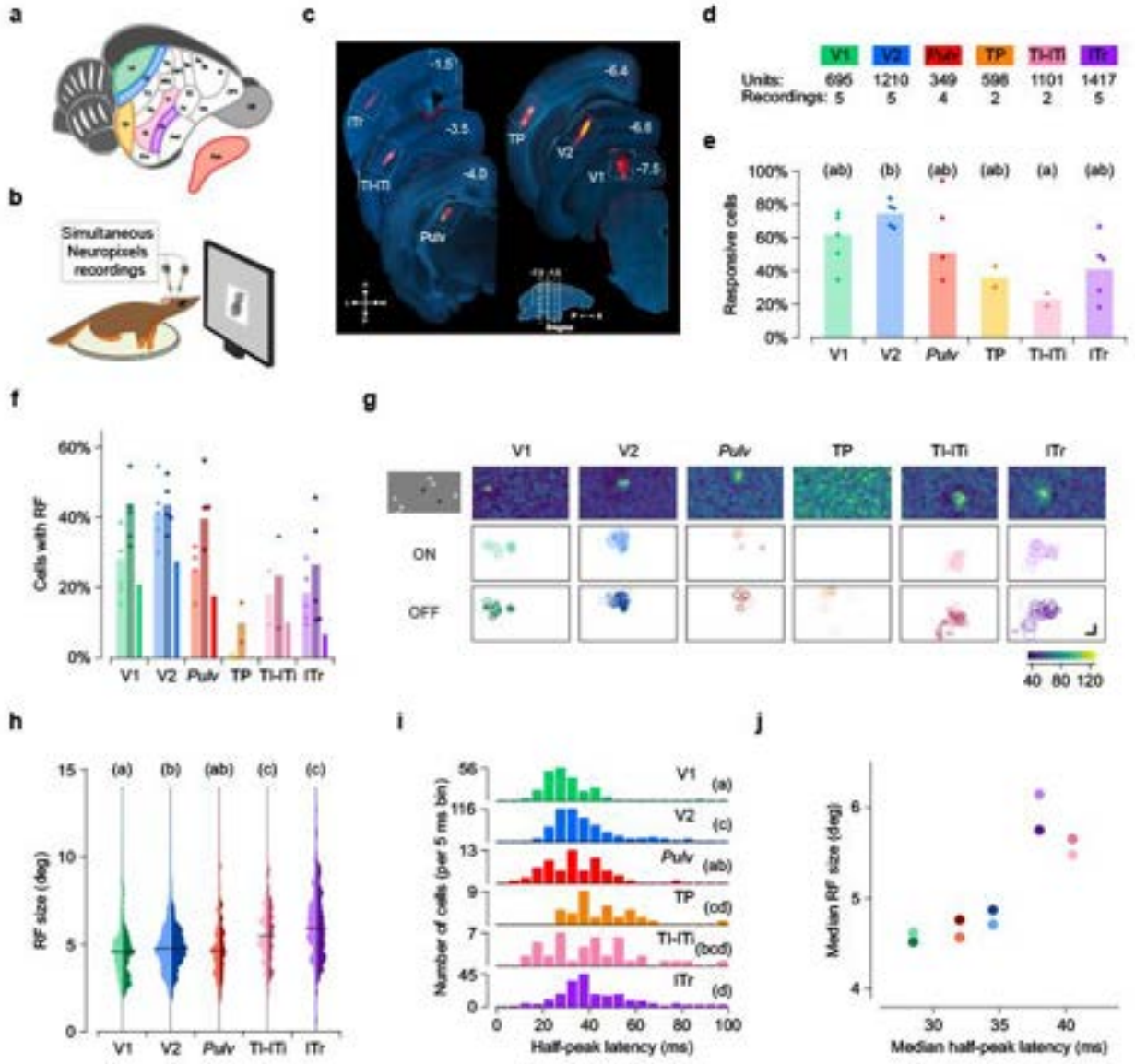

a
b
Simultaneous
Neuropixels
recordings
c
-1.5
-3.5
-4.8
-6.4
-6.6
-7.5
ITr
TI-ITi
Pulv
TP
V2
V1
d
V1 V2 Pulv TP TI-ITi ITr
Units: 695 1210 349 598 1101 1417
Recordings: 5 5 4 2 2 5
e
Responsive cells
100%
80%
60%
40%
20%
0%
(ab) (b) (ab) (ab) (a) (ab)
V1 V2 Pulv TP TI-ITi ITr
f
Cells with RF
60%
40%
20%
0%
V1 V2 Pulv TP TI-ITi ITr
g
V1 V2 Pulv TP TI-ITi ITr
ON
OFF
40 80 120
h
RF size (deg)
15
10
5
0
(a) (b) (ab) (c) (c)
V1 V2 Pulv TI-ITi ITr
i
Number of cells (per 5 ms bin)
56
116
13
9
7
45
0
V1 (a)
V2 (c)
Pulv (ab)
TP (cd)
TI-ITi (bcd)
ITr (d)
0 20 40 60 80 100
Half-peak latency (ms)
j
Median RF size (deg)
6
5
4
30 35 40
Median half-peak latency (ms)

Figure 1: High-throughput electrophysiological recordings along the tree shrew visual pathway reveal a functional hierarchy. (a) Schematic of a tree shrew brain. Color key for recorded areas applies to all figures. **(b)** Schematic of tree shrew head-fixed electrophysiological recording with Neuropixels probes. **(c)** Coronal sections showing electrode traces marked with DiI (red) in each targeted area (white outlined boxes). Numbers indicate rostrocaudal position relative to Bregma (see inset). **(d)** Number of recordings per area and the total number of units identified across all recordings in each area. **(e)** Percentage of visually responsive cells in each area, i.e., cells that responded with elevated average firing rates to at least one of the visual stimuli we tested (static gratings, naturalistic textures and noise, and object stimuli). Dots indicate individual recordings, bars indicate averages across recordings. Letters indicate Tukey grouping. Tukey analysis ($\alpha = 0.05$) after ANOVA, $F_{5,17} = 4.0$, $p < 0.014$). **(f)** Percentage of visually responsive units (cf. **Fig 1e**) exhibiting receptive fields (RFs) for each of the six recorded areas. Left (lighter, ON), center (darker, OFF) and right (ON/OFF) bars for each area. Dots represent results from individual recording sessions. **(g)** Distribution of RF locations across the visual field. Top row: RF maps for example units, one per area. Middle and bottom rows: position and sizes of all ON and OFF RFs (respectively) in a representative recording from each area. Shading indicates RF quality (*Methods*). Each white box represents ±54° horizontally and ±38° vertically. Scale bar: 15°. Top left: one frame of sparse noise stimulus used to map RFs. **(h)** Distribution of ON (left, lighter) and OFF (right, darker) RF sizes for each area. Letters indicate Tukey grouping. Tukey analysis ($\alpha = 0.05$) after ANOVA, $F4,1532 = 36.7$, $p < 10-28$; TP was excluded from this analysis because of the very low number of cells with receptive fields in this area. **(i)** Histogram of the latencies to half-peak response in visually responsive cells in each area. Tukey analysis ($\alpha = 0.05$) after ANOVA $F_{5,1145} = 20.2$, $p < 10^{-18}$. **(j)** Comparison of the hierarchy inferred from RF size (y-axis) with that inferred from response latency (x-axis). Each dot represents the median of the data for a given area (hue), with ON and OFF RFs represented by light and dark dots respectively.

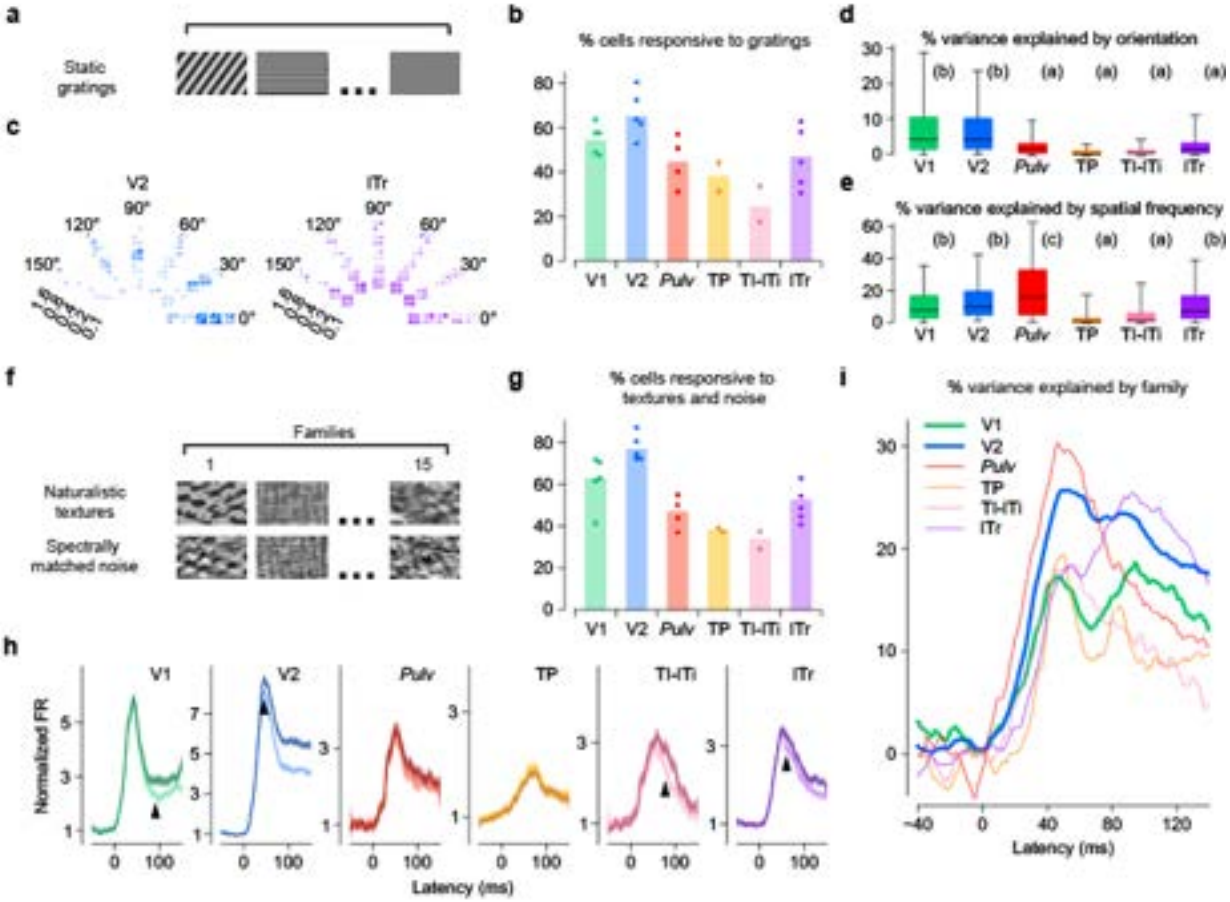

a
Static gratings
b
% cells responsive to gratings
V1 V2 Pulv TP TI-ITi ITr
c
V2
ITr
90°
120°
60°
150°
30°
0°
d
% variance explained by orientation
(b) (b) (a) (a) (a) (a)
e
% variance explained by spatial frequency
(b) (b) (c) (a) (a) (b)
f
Families
1
15
Naturalistic textures
Spectrally matched noise
g
% cells responsive to textures and noise
h
Normalized FR
Latency (ms)
i
% variance explained by family
Latency (ms)

Figure 2: Orientation and spatial frequency of static gratings are first encoded in V1 while textures and noise stimuli are first differentiated in V2. (a) Example frames of static grating stimuli. Stimuli were varied in orientation, spatial frequency (SF), and phase, and were interleaved with gray frames. **(b)** Percentage of visually responsive cells (cf. **Fig 1e**) that responded to static gratings in individual recording sessions (dots) and averaged across recording sessions (bars). **(c)** Responses of a representative V2 and ITr cell to static gratings differing in orientation (represented circumferentially), SF (represented radially; cycles/°), and phase (four small quadrants). Each dot represents a single trial; color intensity represents responses strength. **(d)** Percentage of variance of individual cells' responses explained by orientation of the stimulus. Boxes represent 25th, 50th and 75th percentile; whiskers 5th and 95th. Letters indicate Tukey grouping. Tukey analysis ($\alpha = 0.05$) after ANOVA, $F_{5,1099} = 25.9$, $p < 10^{-24}$. **(e)** Same for spatial frequency. Tukey analysis ($\alpha = 0.05$) after ANOVA, $F_{5,1099} = 19.4$, $p < 10^{-17}$. **(f)** Example frames of naturalistic texture (top) and spectrally matched noise (bottom). **(g)** Percentage of visually responsive cells (cf. **Fig 1e**) that responded to naturalistic texture or spectrally matched noise stimuli in individual recording sessions (dots) and averaged across recording sessions (bars). **(h)** Time courses of population responses in each area to naturalistic texture (darker lines) and spectrally matched noise (lighter lines). Black arrows indicate the latency at which the two curves first significantly differed from each other (Two-tailed t-test, $p < 0.01$). Shaded areas are standard errors of averages across cells. **(i)** Percentage of variance in neural activity explained by texture image family (15 classes, cf. (f)).

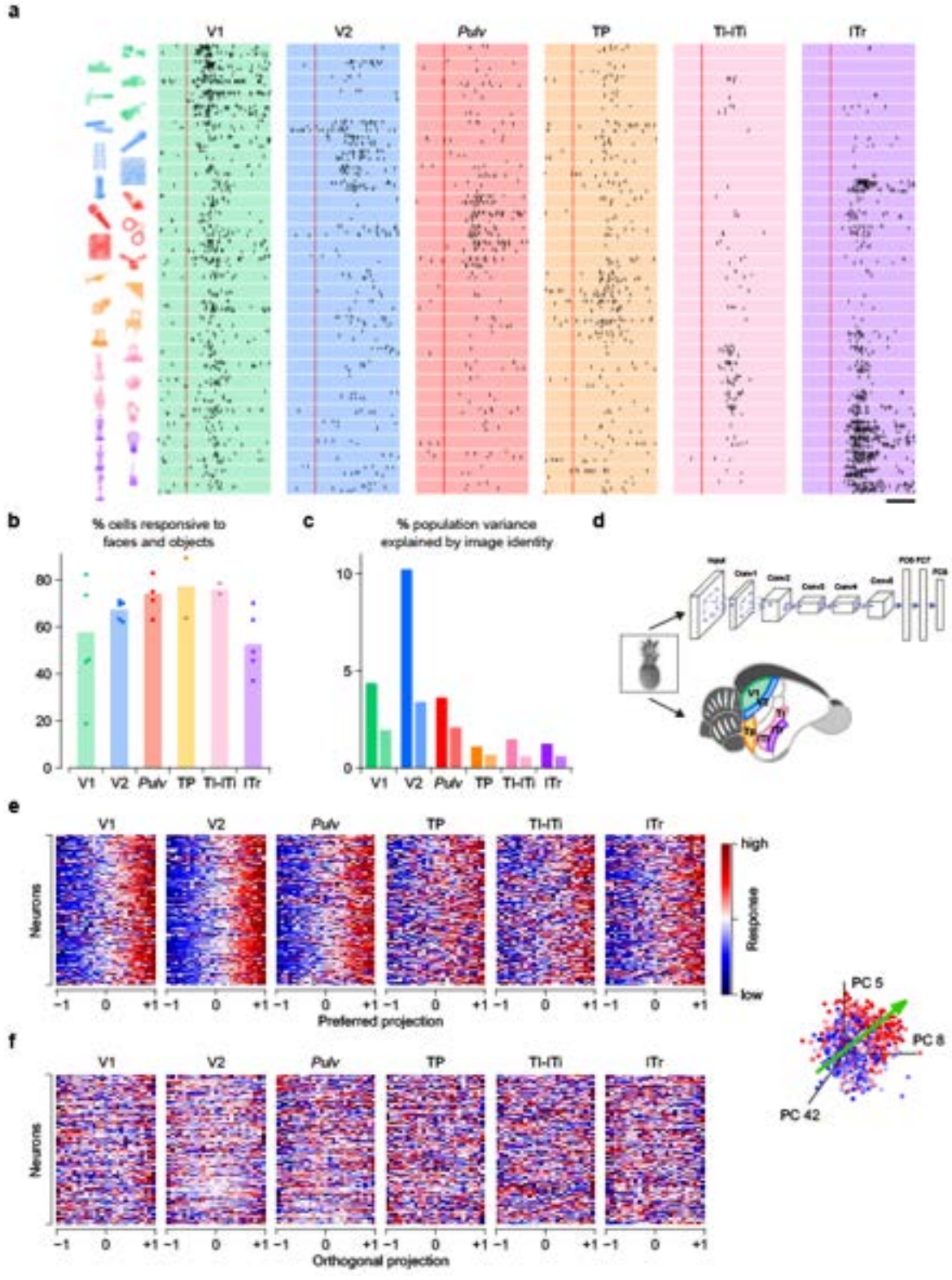

a
V1
V2
Pulv
TP
TI-ITI
ITr
b
% cells responsive to faces and objects
80
60
40
20
0
V1 V2 Pulv TP TI-ITI ITr
c
% population variance explained by image identity
10
5
0
V1 V2 Pulv TP TI-ITI ITr
d
e
V1
V2
Pulv
TP
TI-ITI
ITr
Neurons
high
Response
low
−1 0 +1
Preferred projection
PC 5
PC 8
PC 42
f
V1
V2
Pulv
TP
TI-ITI
ITr
Neurons
−1 0 +1
Orthogonal projection

Figure 3: Objects are encoded across all visual areas through axis coding. (a) Spike raster plots for representative visually-active cells from each of the areas in response to six groups of object stimuli, each optimal for one of the cells (stimuli shown on the left). Each dot represents an action potential in one of up to 10 presentations of the stimulus; red line indicates stimulus onset. Scale bar: 50 ms. **(b)** Percentage of visually responsive cells (cf. **Fig 1e**) that responded to object stimuli in individual recording sessions (dots) and averaged across recording sessions (bars). **(c)** Percentage of variance of neural responses explained by object stimulus identity in each area (left bars) and the amount of variance that can be explained by low level feature image indices (right bars). **(d)** Schematic illustrating the processing of visual stimuli in layers of the artificial neural network AlexNet (top) and in areas of the tree shrew ventral visual pathway (bottom). **(e)** Normalized neural responses to object images for 100 randomly selected cells in each of the six areas as a function of position of that image along the given neuron's preferred axis in AlexNet FC6 space ("object space"). The x-axis is rescaled so that the range [–1,1] covers 98% of the stimuli. Inset: Preferred axis (green arrow, *Methods*) of a representative cell (area V2) in object space. The coordinate axes represent the three AlexNet principal components (PCs) that most align with the cell's preferred axis. Each dot represents an image, color coded by the strength of the cell's response to that image (blue: low, red: high). **(f)** Responses as a function of normalized position along each cell's principal orthogonal axis, i.e., the axis in object space orthogonal to the neuron's preferred axis that captured the most variance in AlexNet activations (*Methods*).

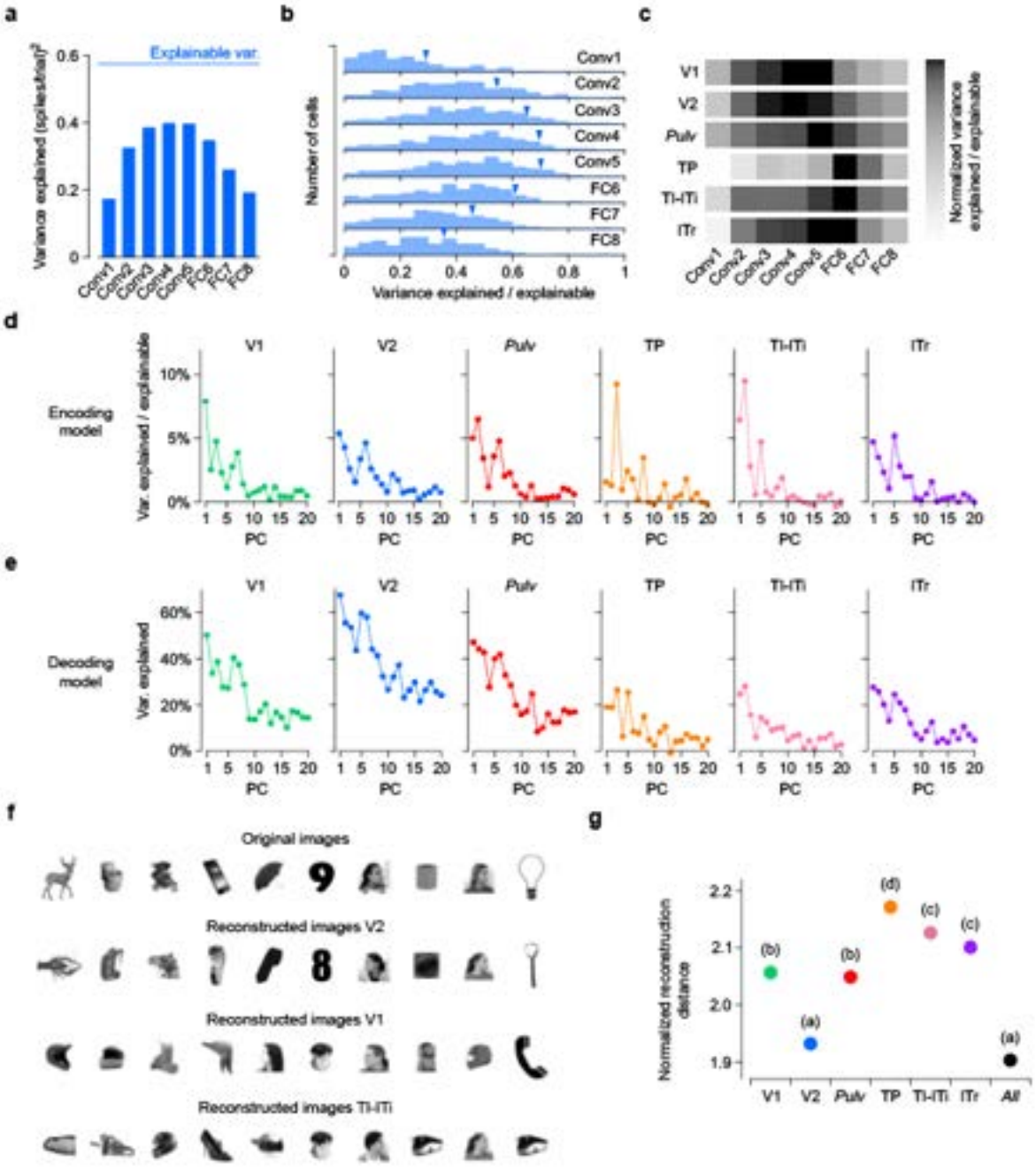

a
b
c
Explainable var.
Variance explained (spiesd/sd)²
Conv1
Conv2
Conv3
Conv4
Conv5
FC6
FC7
FC8
Number of cells
Variance explained / explainable
V1
V2
Pulv
TP
TI-ITi
ITr
Normalized variance explained / explainable
d
e
Encoding model
Decoding model
Var. explained / explainable
Var. explained
PC
f
g
Original images
Reconstructed images V2
Reconstructed images V1
Reconstructed images TI-ITi
Normalized reconstruction distance
All

Figure 4: Neural representation of object stimuli in tree shrew ventral areas reveals optimal feature decoding in area V2. (a) Variance of the responses of a representative V2 cell explained by individual AlexNet layers. Blue line: explainable variance of the cell. **(b)** Histograms of explained variance by different layers of AlexNet for responses of responsive cells (*n* = 602) in area V2. Blue triangles mark values for the cell from (a). **(c)** Normalized explained variance by AlexNet layers for each tree shrew visual area (*Methods*). **(d)** Variance of encoded neural activity in different areas explained by individual AlexNet FC6 principal components (PCs) as a percentage of explainable variance in that area. **(e)** Percentage of variance of AlexNet FC6 features that can be explained by decoding from the neural responses in different areas. **(f)** Ten examples of original images presented to the tree shrew and the images reconstructed from V2, V1 and TI-ITi, i.e., closest images to the predicted responses from AlexNet FC6 from an auxiliary database of images that were not shown to the animal (*Methods*). **(g)** Average decoding distance for each tree shrew visual area between AlexNet FC6 activations predicted from neural activity and actual activations for each image, normalized by theoretical best decoding distance (*Methods*).

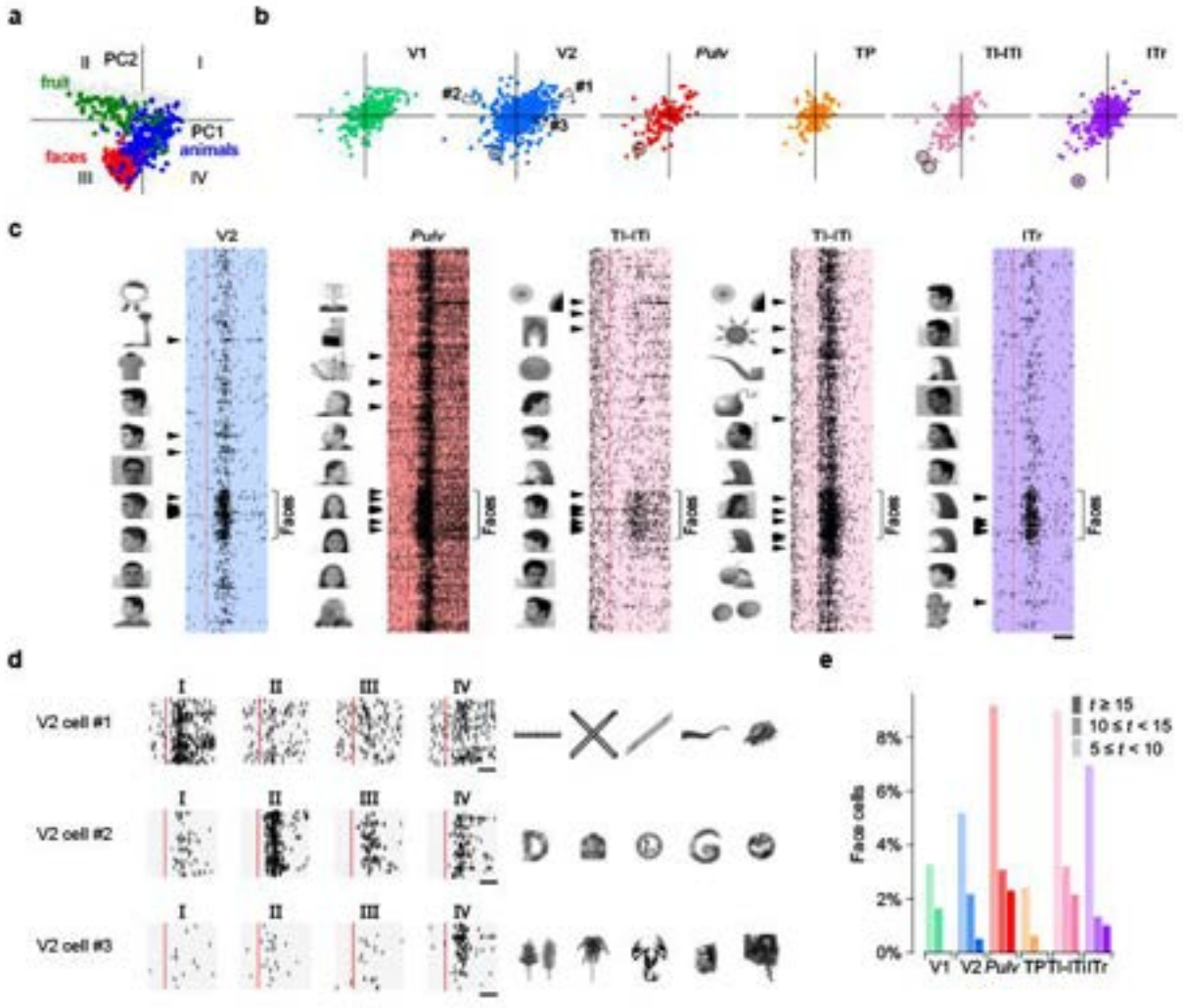

a
b
c
d
e
PC1
PC2
fruit
faces
animals
V1
V2
Pulv
TP
TI-ITI
ITr
Faces
V2 cell #1
V2 cell #2
V2 cell #3
Face cells
8%
6%
4%
2%
0%
V1 V2 Pulv TP TI-ITI ITr

Figure 5: Single cells across the tree shrew ventral stream show selectivity for different sectors of object space including faces. (a) Projections of 1593 object images onto object space (the first two PCs from AlexNet layer FC6) with images from several categories (faces, animals, fruits) indicated. **(b)** Projections of the preferred axes of all cells onto object space. **(c)** Raster plots of several representative face-selective cells (circled in (b)) responding to face and object stimuli. The 10 most preferred images for each cell are shown to the left of each raster. Arrowheads mark responses to those images. Red line: stimulus onset. Scale bar: 5 ms. **(d)** Raster plots of three representative V2 cells (arrowheads in (b)) with preferred axes in quadrants I, II, and IV. Twenty stimuli from each quadrant were randomly chosen to generate raster plots. Scale bar: 50 ms. Right: Top five preferred images for each cell. **(e)** Histograms of t-scores for face selectivity across areas.

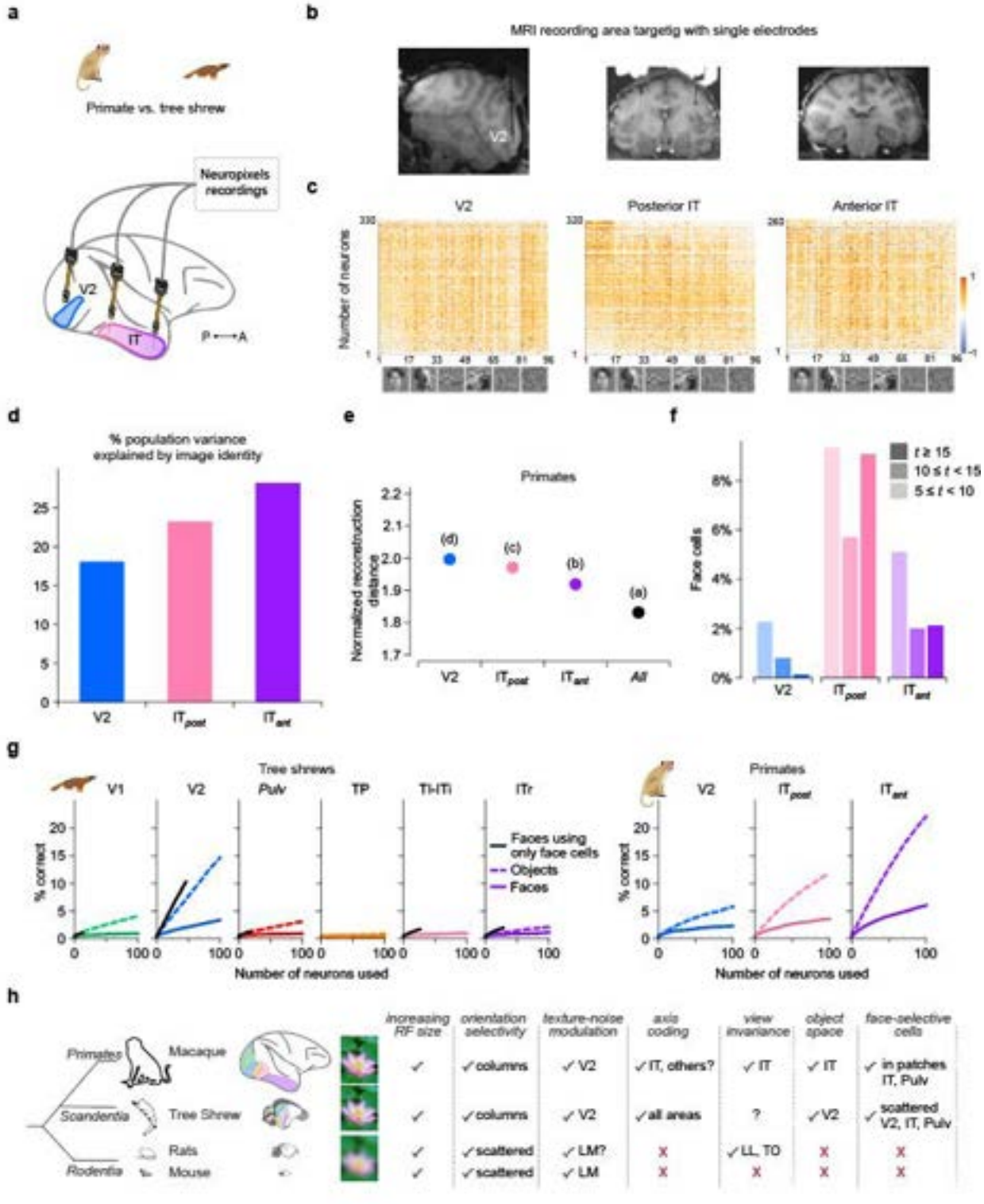
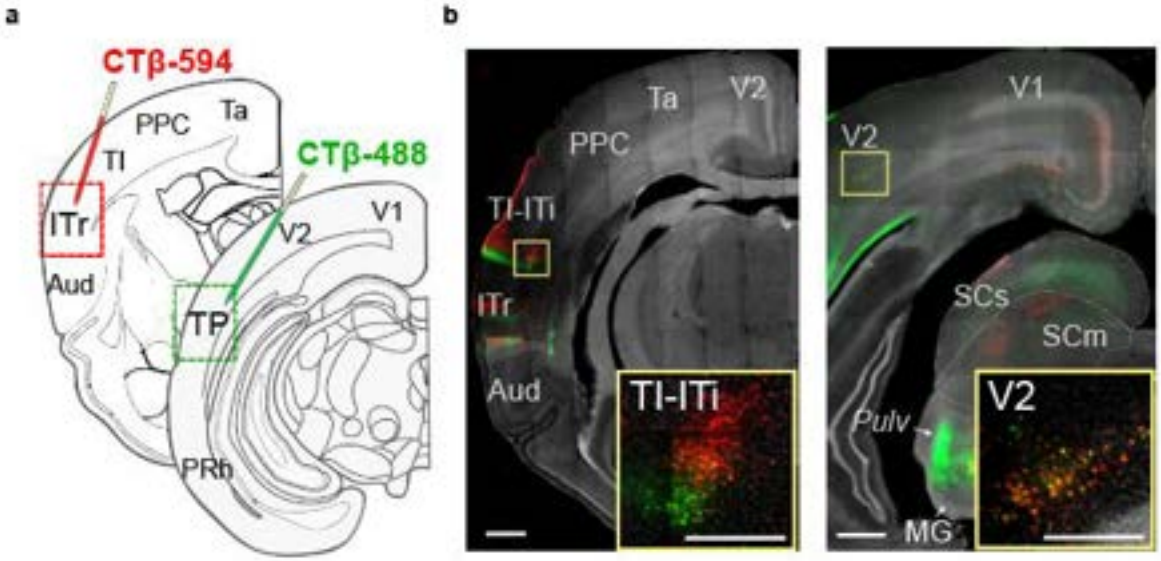

a
Primate vs. tree shrew
Neuropixels recordings
V2
IT
P ⟷ A
b
MRI recording area targetig with single electrodes
V2
c
V2
Posterior IT
Anterior IT
Number of neurons
d
% population variance explained by image identity
V2
IT post
IT ant
e
Primates
Normalized reconstruction distance
(d)
(c)
(b)
(a)
V2
IT post
IT ant
All
f
t ≥ 15
10 ≤ t < 15
5 ≤ t < 10
Face cells
8%
6%
4%
2%
0%
V2
IT post
IT ant
g
Tree shrews
V1
V2
Pulv
TP
TI-ITi
ITr
Faces using only face cells
Objects
Faces
% correct
Number of neurons used
Primates
V2
IT post
IT ant
% correct
Number of neurons used
h
Increasing RF size
orientation selectivity
texture-noise modulation
axis coding
view invariance
object space
face-selective cells
Primates
Macaque
✓
✓ columns
✓ V2
✓ IT, others?
✓ IT
✓ IT
✓ in patches IT, Pulv
Scandentia
Tree Shrew
✓
✓ columns
✓ V2
✓ all areas
?
✓ V2
✓ scattered V2, IT, Pulv
Rodentia
Rats
✓
✓ scattered
✓ LM?
X
✓ LL, TO
X
X
Mouse
✓
✓ scattered
✓ LM
X
X
X
X

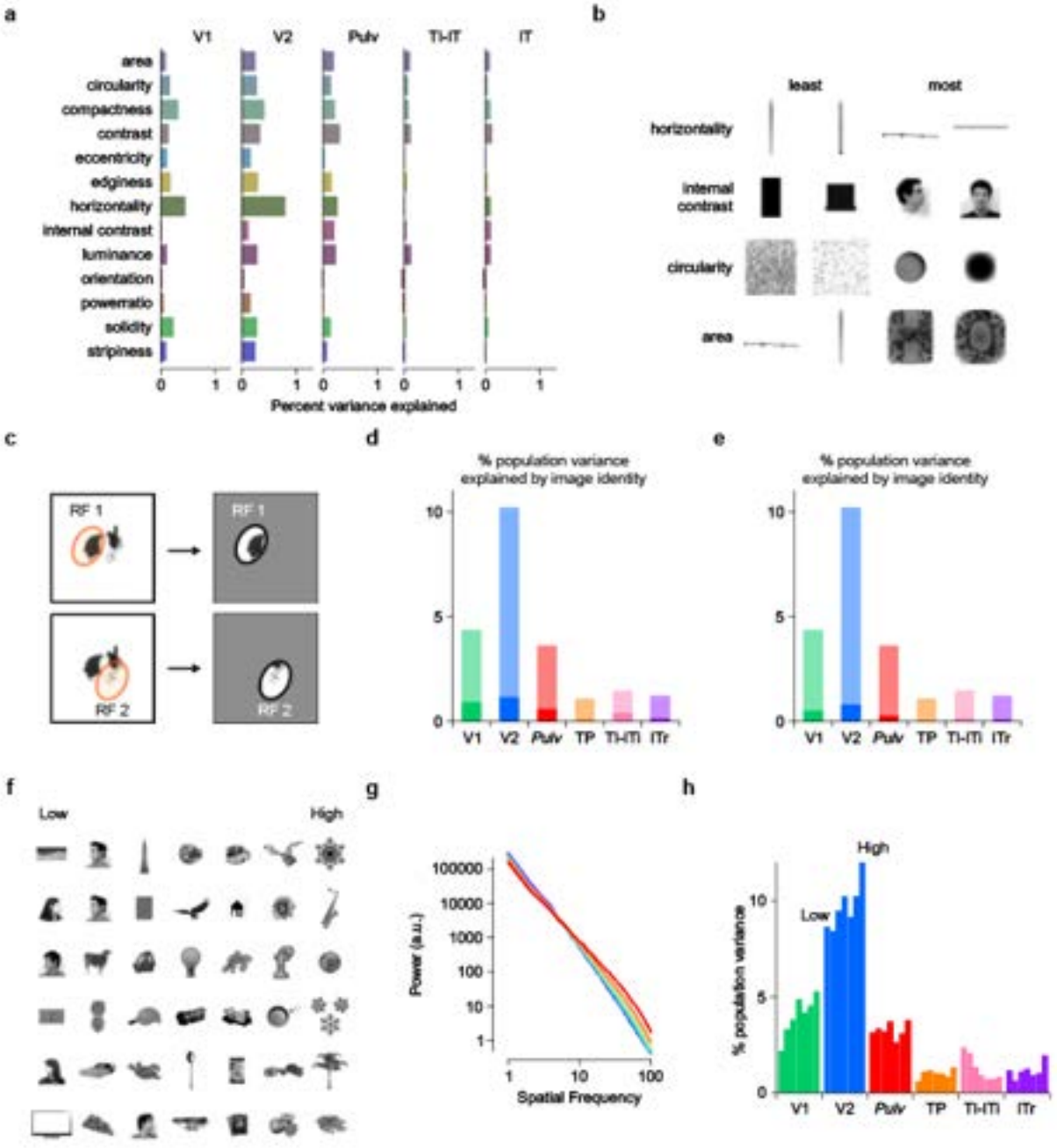
Figure 6: Comparison of object responses between primate and tree shrew ventral stream **(a**) Schematic of recordings in primate. **(b)** Simultaneous Neuropixels recordings from three nodes in macaque monkey cortex. Neuropixels NHP 1.0 probes were inserted into V2, posterior IT and anterior IT cortex.: (**c**) Responses of 330, 320 and 260 cells in V2, posterior IT and anterior IT, respectively (rows), to 96 stimuli composed of faces and objects (columns). Only visually responsive cells were included (t-test, $p<0.05$). (**d**) Percentage of variance of neural responses explained by object stimulus identity in each area. (**h**) Average decoding distance for each visual area between AlexNet FC6 activations predicted from neural activity and actual FC6 activations for each image, normalized by theoretical best decoding distance (*Methods*). (**f**) Histograms of t-scores for face selectivity across areas. **(g)** Decoding performance for individual object identity (dashed lines) or face identity (solid lines) as a function of number of cells used by the classifier. Note the overlap of the two lines for TI-ITi. Black lines indicate decoding performance for face identity using only face cells (t-score greater than 5). Dashed gray lines: chance level for object decoding. (**h**) Schematic comparing macaque, tree shrew, and rodent visual systems.

Extended Data Fig. 1



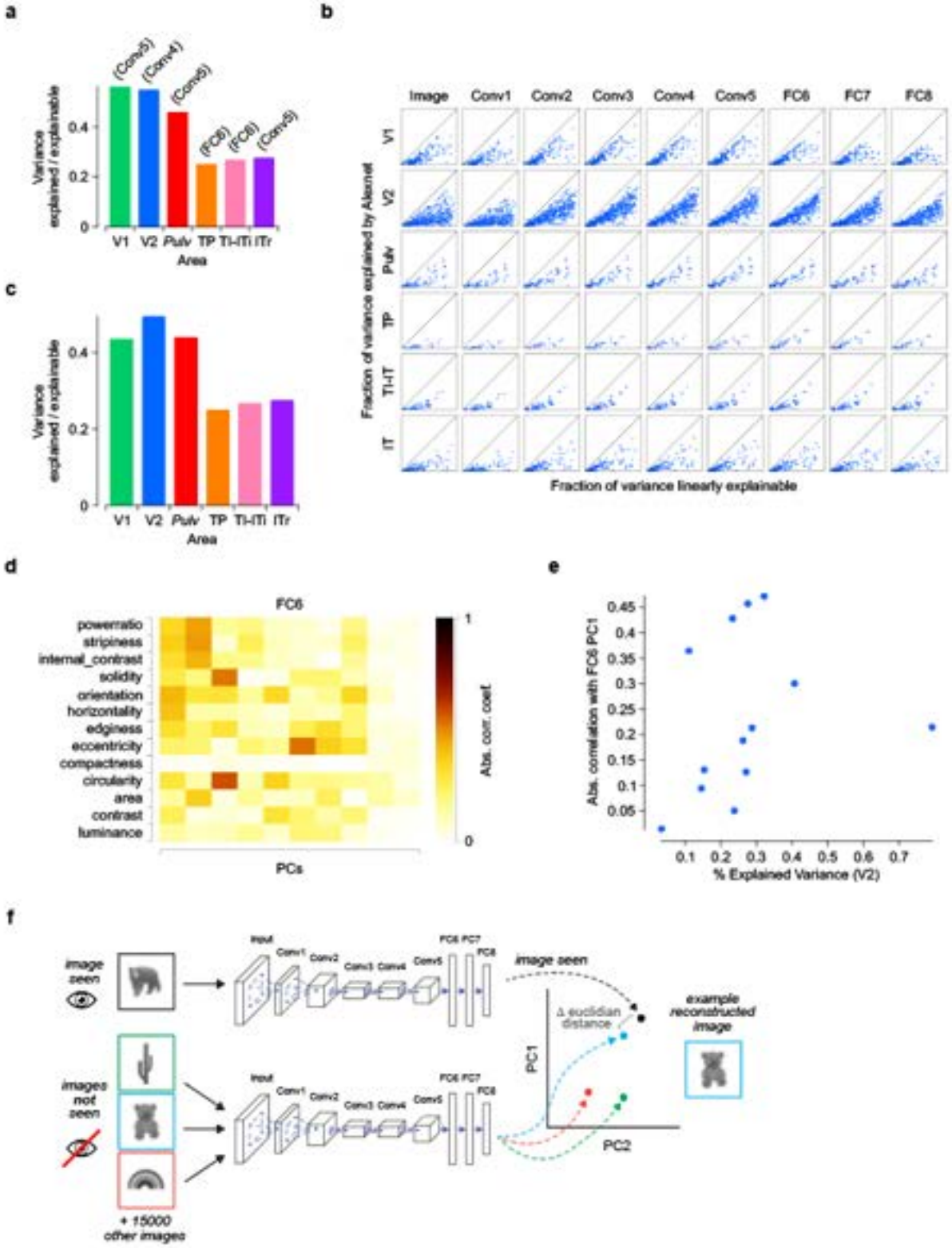
Extended Data Figure 1: Anatomical inputs to intermediate (TP) anterior (ITr) nodes of the tree shrew ventral pathway. (a) Schematic of injections of retrograde tracer CTβ-488 (green) into TP and CTβ-594 (red) into ITr. **(b)**Two coronal histological sections showing retrogradely labeled cells projecting to TP (green) and ITr (red) and counterstained with DAPI (grey). Scale bars: 1 mm / 0.5 mm (insets).

Extended Data Fig. 2



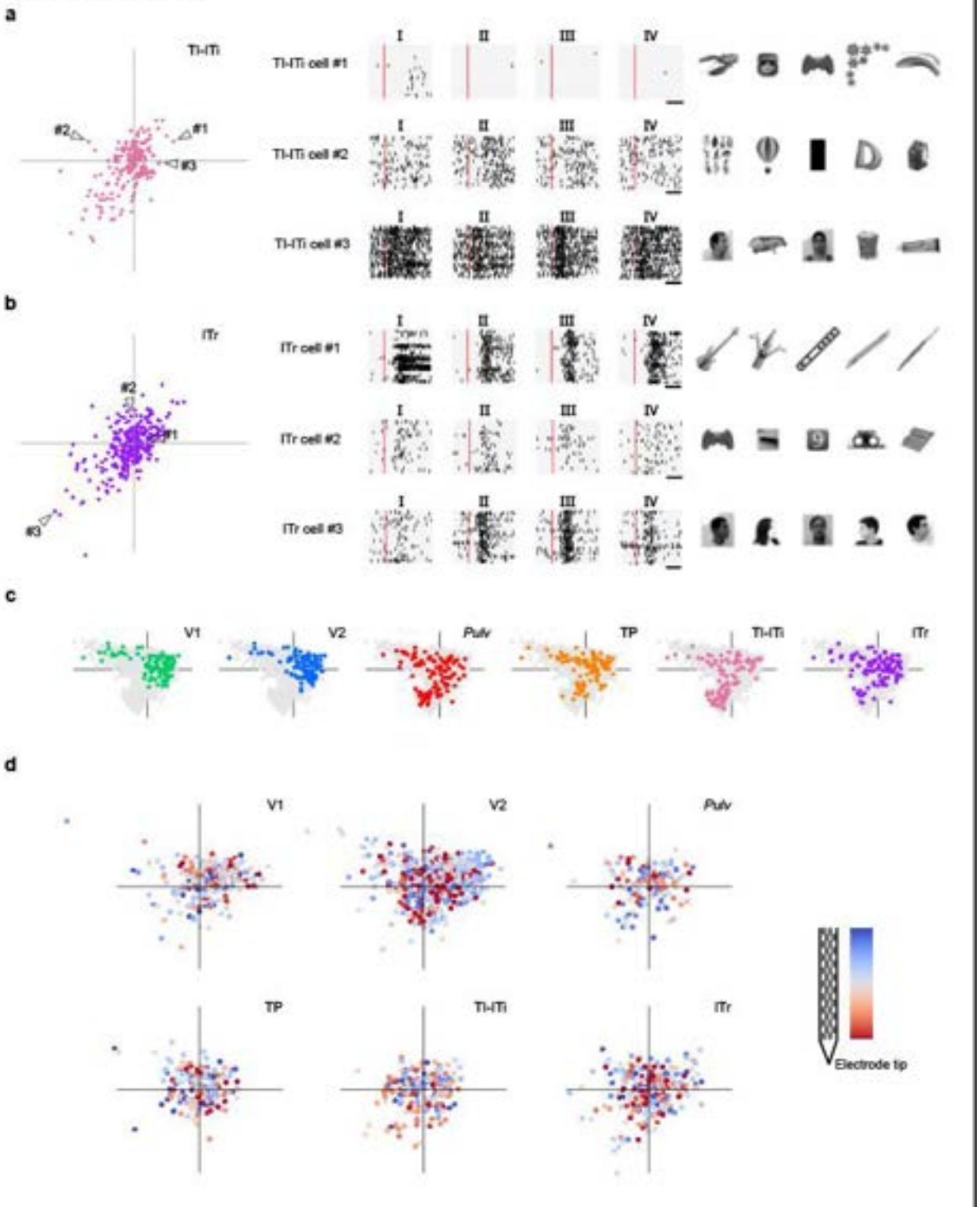
Extended Data Figure 2: Object responses are largely not accounted by low-level features **(a)** Histogram indicating the average fraction of variance in the firing rate explained by various low-level image feature indices. **(b)** Examples of the 3 images with the lowest (left) and highest (right) value for horizontality, internal contrast, circularity and area. **(c)** Schematic of quantification of luminance and contrast impinging on each receptive field. We computed the average luminance and contrast (second derivative of luminance) falling inside by the ON and OFF receptive fields of each cell, and average across the two. **(d)** Percentage of variance of neural responses explained by object stimulus identity in each area. Dark bars correspond to the part of the variance accounted for by luminance impinging each receptive field. **(e)** Same, but dark bars correspond to contrast. **(f)** Representative objects with increasing high spatial frequency content from low (leftmost column) to high (rightmost column). **(g)** Power spectrum across groups of images in (a). **(h)** Percentage of variance of neural responses explained by object stimulus identity in each area, separated into categories based on spatial frequency.

Extended Data Fig. 3



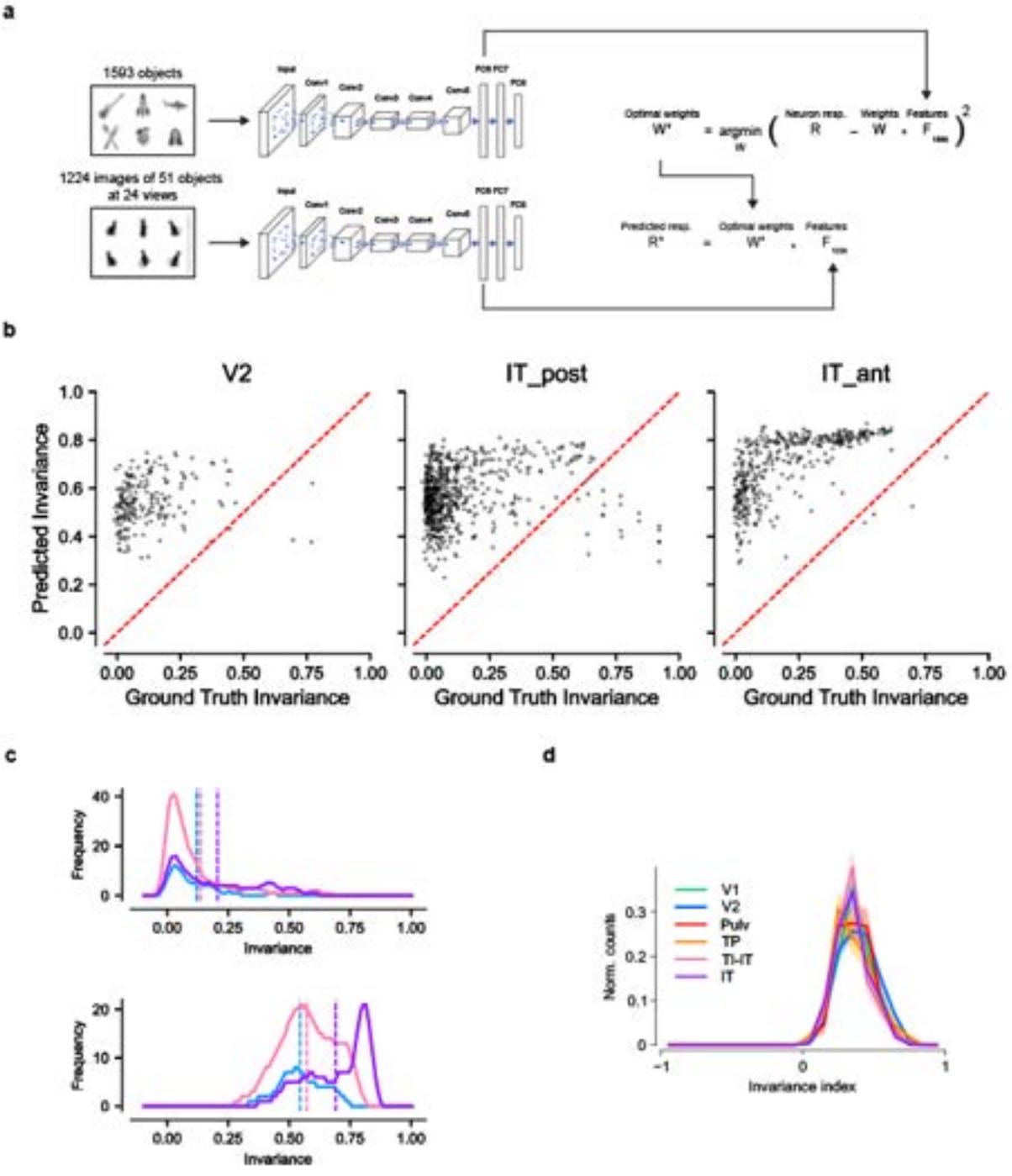
Extended Data Figure 3: Representation of different sectors of object space by cells in each tree shrew visual area. (a) Aggregate explanatory power of the AlexNet layer that best explained each given area. **(b)** Fraction of variance in the firing rates of individual cells (*dots*) explained by different AlexNet layers plotted against the fraction of the total explainable variance in that cell (*Methods*). **(c)** Aggregate explanatory power of AlexNet layer FC6 over different areas. **(d)** Absolute correlation between image indexes and the first 10 PCs of AlexNet FC6. **(e)** Scatter plot between the percentage of explained variance by a given feature in V2 neural responses and the absolute correlation of that feature with FC6 PC1 activations in AlexNet. **(f)** Schematic of image reconstruction.

Extended Data Fig. 4



Extended Data Figure 4: Cells selective to different sectors of object space with no obvious topographical organization in object space for each area. (a) Projections of each TI-ITi cell's preferred axis onto the first two PCs of object space (replicated from **Fig. 5b**). Right: Raster plots of three representative TI-ITi cells from quadrants I, II, and IV indicated by letters; twenty stimuli from each quadrant were randomly chosen to generate raster plots. Scale bar: 50 ms. Top five preferred images for each cell. **(b)** Same for ITr. **(c)** All 1593 images projected onto the first two PCs of object space. Colored dots: the 100 images that elicited the strongest responses overall in each area. **(d)** Selectivity of cells in each area as a function of recording depth along the Neuropixels probe. In each of the six plots, each dot represents one cell, the color of the dots indicates the depth at which the cell was recorded (inset, right), and the position of the dot indicates the mean projection of the 10 most preferred images onto the first two PCs of object space.

Extended Data Fig. 5



Extended Data Figure 5: DNN-predicted indices of view invariance are equally high across all tree shrew visual areas. (a) Schematic showing workflow for predicting neuron responses for a new set of stimuli. 1593 images were passed through AlexNet (top). Activations in AlexNet layer FC6 were used to linearly predict neural responses evoked by each image when shown to the animal. This yields a weight matrix W that optimally predicts a neuron's response based on the image features F. Next, the weight matrix is used to predict neuron responses to 1224 images consisting of 51 objects at 24 views that were not shown to the tree shrew (bottom). **(b)** Correlations of the ground truth and predicted invariance across cells for primate area V2, posterior IT, and anterior IT. **(c)** Histograms of invariance indices (*Methods*) of macaque posterior and anterior IT neurons, calculated from actual responses (top) and predicted responses (bottom). Shading indicates bootstrap resampling; vertical lines indicate medians. **(d)** Histograms of invariance indices of predicted responses across all tree shrew areas.

Chapter III

Probing feedforward and feedback pathways in the tree shrew visual cortex with electrical microstimulation

Abstract:
Visual processing occurs along hierarchically organized areas that are reciprocally connected, with both feedforward and feedback circuits. These interactions allow higher-order areas to modulate and refine sensory input while enabling lower-order areas to influence complex visual representations. To investigate the functional properties and connectivity of feedback and feedforward neurons, we combined high-density electrophysiological recordings with electrical microstimulation in the tree shrew visual cortex. By selectively stimulating primary visual cortex (V1) and recording from secondary visual cortex (V2), we identified distinct classes of V2 neurons, including feedforward and feedback neurons, based on the directionality of their projections and studied their response properties. Our preliminary functional analyses did not reveal any clear distinction between feedforward and feedback neurons. Contrary to the idea that feedback neurons transmit specialized feature-specific signals[1–4], we found that V2 feedback neurons carry diverse visual information on par with other V2 neurons. Furthermore, both feedforward and feedback cells also carry information about perceptual conflict that arises when distinct images are presented to each eye. However, feedback neurons did exhibit few distinct properties, including smaller receptive fields and spatial offsets relative to their V1 inputs, potentially enabling precise and localized modulation of V1 activity. Overall, these results establish electrical stimulation combined with Neuropixels probe recordings as a useful method for distinguishing cell classes.

INTRODUCTION

The unique neural computations performed by visual neurons are largely shaped by their inputs that convey information about the outside world[5]. These inputs are received, processed, and relayed through neural projections that influence their downstream targets. Much work has focused on how information flows along the visual hierarchy in a feedforward direction to construct representations of shapes, color, and form[6]. However, understanding the computations performed along each node in the hierarchy must be considered in the context of its interactions with other nodes, not just its individual processing.

Tree shrew area V2 is thought to play a major role in form vision as it sits at the gateway to the ventral visual stream[7]. Previous studies have characterized some of the functional properties of cells in this area, including sensitivity to orientation, binocularity, texture, and form[8–10]. However, how the cells recorded in these studies fit into the broader hierarchical network remains unclear, especially since the directionality of their projections, whether they provide feedback to earlier areas (such as V1) or feedforward input to later areas, was not explicitly measured.

In tree shrew and other highly visual mammals, area V2 receives most of its cortical input from V1, while also sending strong feedback projections back to V1. Investigating the visual processes arising from the interactions between these two areas is challenging, as extracellular recordings from a single area cannot determine the target of projection neurons. Feedback connections at many stages of the visual hierarchy vastly outnumber their feedforward connections[11–13], yet their functions remain poorly understood, especially in relation to the feedforward projections. Several theories have been proposed to explain the role of feedback in the visual system, including its involvement in contextual modulation, error correction, and the integration of information across spatial and temporal domains[14]. Central to distinguishing between these theories is the question

whether feedback projections convey functionally specific signals that refine particular aspects of visual processing or whether they carry more generalized signals that mirror the diversity of feedforward information.

Here, we address this question by systematically examining the response properties of V2 cells based on the directionality of their connection with V1 in the tree shrew visual system. Using a combination of electrical stimulation and functional characterization, we identified and recorded from neurons in V2 that receive inputs from V1 (i.e. feedforward recipient) or project to V1 (i.e. feedback). We then compared the responses across the different classes of V2 cells to visual stimuli, including static gratings and naturalistic textures. We found that feedback neurons exhibited response properties similar to that of other V2 neurons. This finding suggests that feedback neurons appear to carry the same diverse visual information processed by other neurons in V2. V2 feedback cells did differ with respect to their spatial properties, including smaller receptive field sizes and their offset relative to V1. When presenting conflicting naturalistic textures to each eye, most V2 cells including feedforward and feedback reduced their responses relative to non-conflict conditions. This may suggest that through feedback connections V2 they could adjust its V1 inputs in conditions relevant for binocular disparity or visual field mismatch, potentially aiding in error correction.

RESULTS

Identification of feedforward and feedback neurons in the tree shrew visual cortex

To understand vision, it is not only critical to know what signals V1 sends to V2, but also how that information is transformed and sent back to modify V1. To address this, we sought to functionally classify neurons in V2 based on their connections with V1. Combining high–channel count Neuropixels probes[15] with bipolar electrical microstimulation in the tree shrew visual cortex (**Supplemental Figure 1**) allowed for selective stimulation of individual neurons extracellularly to identify and characterize the functional properties of V2 neurons and the directionality of their axonal projections. We electrically stimulated the primary visual cortex (V1) while simultaneously recording in the secondary visual area (V2) to disentangle neurons sending feedback signals (i.e. antidromic) from those receiving feedforward signals (i.e. orthodromic; **Figure 1A**). Electrodes were anatomically mapped at the end of each experiment to confirm the location of both the stimulation site in V1 and the Neuropixels recording site in V2 (**Figure 1B**). Where electrical microstimulation induced an electrical artifact that interfered with V2 recordings, this artifact was removed *ex post* using the SALPA algorithm[16] (**Supplemental Figure 2**; Methods).

To characterize neurons as antidromic (i.e. providing feedback connections to V1), we relied on several metrics to determine the direction of the axonal projection for recorded V2 cells[17,18]. We first performed a ramp test by delivering biphasic pulses of increasing current amplitude. Based on the ramp test, the antidromic responses were identified using two defining features: (1) the temporal jitter of the responses was below 150 µs (**Figure 1C**, top right inset), well below the timing precision of synaptic transmission; and (2) responses to pulses near the activation threshold were bimodal, i.e., in each individual trial, the response was either fully present and equal to the response at higher currents, or entirely absent (**Figure 1C**, bottom right inset), in opposition to electrical artifacts, which would be expected to scale with stimulus current and be present equally in all trials in which the same current was applied.

The second test we performed to determine the directionality of V2 projection cells was a collision test[19]. With this test, the antidromic nature of a response could be confirmed by the occasional occurrence of transmission failure due to action potential collision (**Figure 1D**, Methods). If cells met all of our defined criteria in both the ramp and collision test, they were defined as “confirmed

antidromic" for all further analyses. For cells not identified as antidromic, we classified cells as orthodromic (i.e. receiving feedforward signals from V1) if the responses adhered to a timing precision of better than 2.5 ms that occurred in at least 20% of trials, on par with the timing of one synaptic transmission.

While this criterion was suitable for cells that had a sufficiently high spontaneous firing rate in which the potential for a collision is highly likely, it could limit the classification of true antidromic cells with low firing rates. To facilitate the classification of these cells, we applied a free split/merge expectation maximization (FSMEM) algorithm that fits a mixture of gaussian onto the responses of a neuron to electrical stimulation (**Figure 1E**). Based on latency and width of all peaks obtained using the FSMEM algorithm classification, we found that antidromic and orthodromic neurons distinctly clustered into two separate groups (**Figure 1F**). Orthodromic neurons exhibited spikes with latencies and widths typical of feedforward processing, while antidromic neurons showed distinct characteristics consistent with feedback signaling. Notably, the cells that qualified as potential antidromic based on the above metrics except that they did not have a confirmed collision test (i.e. "putative antidromic") clustered together with the confirmed antidromic cells. These putative antidromic cells are low firing and, therefore, may not have an observable collision nor do they fail the collision test. The clustering of both confirmed and putative antidromic cells further indicate that these are feedback cell projecting back from V2 to V1. In total, our sample comprised of 397 orthodromic and 64 antidromic neurons across ten separate recordings (**Figure 1G**). Further, we characterized the spike waveforms to ensure that all classified cells were indeed V2 cells and not recorded V1 axons located in the V2 (**Supplemental Figure 2F**).

Feedback cells have small receptive field sizes and fast response latencies

Having characterized the directionality of a subset of V2 cells as antidromic and orthodromic, we asked about the characteristics of the unclassified remaining cells. Do these cells exhibit differential responses to electrical microstimulation that may suggest they form functional classes within the canonical V2 cortical circuity? We identified two additional classes, one that increased (i.e. excited) and the other that decreased their firing rate following microstimulation at a latency suggestive of multi-synaptic connections (i.e. inhibited). Altogether, we identified four classes of V2 cells: antidromic, orthodromic, excited, and inhibited (**Figure 2A-B**). These classes were located in two distinct spatial location along the depth of the probe; the antidromic and inhibited were shallower while the excited and orthodromic were located deeper in the cortex (**Figure 2C**).

We next asked how these identified classes of cells responded to visually stimuli in the absence of microstimulation. All classes, except for the inhibited cells, were visually driven as they displayed a significant increase in their firing rate to visual stimuli as compared to darkness (i.e. spontaneous; **Figure 2D**). To be able to compare V2 responses to that of their V1 inputs, we recorded multiunit activity from the V1 electrode when it was not used for microstimulation. This allowed us to characterize population-level responses of V1 neurons at the site of stimulation by calculating the instantaneous energy (Methods) in V1 to visual stimuli (for e.g. static gratings; **Figure 2E**). This confirmed significant visually driven responses.

We next asked whether the V2 feedback neurons represent a functionally distinct population that differs in their visual response properties from the rest of the V2 population. Do they have distinctive receptive field properties, and how do they map spatially to their V1 inputs? In the absence of electrical stimulation, we mapped the receptive fields of the V1 multiunit activity and V2 neurons in by presenting a locally sparse noise stimulus (Methods; **Figure 2F**). We measured receptive fields of single neurons by fitting a gaussian distribution to the two-dimensional matrix of spike counts at each location of the visual field; ON and OFF receptive fields were computed

separately using responses to white and black squares, respectively. We mapped the V2 neurons receptive field locations relative to the location of the receptive field of V1 multiunit activity (shown in **Figure 2E**). The V2 receptive fields for all classes, including the antidromic, were not completely overlapping with the V1 receptive field. Rather, they were offset relative to the V1 receptive field (examples in **Figure 2G**) as would be expected from previous work[20,21].

A modest majority of visually responsive cells had clear receptive fields. Orthodromic and excited cells were more likely to have RFs than did inhibited cells. Of the antidromic cells, fewer than half had measurable receptive fields (**Figure 2H**). Surprisingly, for antidromic cells that did have receptive fields, those receptive fields were smaller than those of any other class of V2 cells (**Figure 2I**). This result may suggest that feedback cells are likely more similar to that of V1 which may be important for modulating their activity in a highly spatially restricted manner.

Cell class defined by responses to electrical stimulation is predictive of responses to binocular conflict stimuli

Binocularity is an important feature of V2 cells across species[22–25]. This information is thought to arise both from feedforward V1 binocular neurons and other cortical inputs. We therefore asked whether our four classes of cells differ in their binocularity. The tree shrew visual cortex is topographically organized with the binocular zone corresponding to the central visual field[25]. Therefore, we targeted our recording sites to this region as determined by the location of receptive fields in the center of the screen (c.f. **Figures 2F-G**). We then presented images including static grating and textures through custom-made anaglyph goggles that allowed for the independent control of visual input to each eye (**Figure 3A**).

In response to gratings or textures presented to both eyes, we did not observe strong difference in the variance explained across the different classes of V2 cells, with the exception of the inhibited group which was consistently lower (**Figure 3B-C**). This suggests that overall, there was no observable difference in their feature-selectivity. When gratings were presented to only one eye at a time, the vast majority of cells of all cell types responded nearly equally whether those stimuli were presented to the ipsilateral (ipsi) or the contralateral (contra) eye (**Figure 3D**, left “G”). In contrast, when texture stimuli were presented in the same way, the preference for ipsilateral or contralateral presentation was more varied, and was biased toward ipsilateral in all types except the “inhibited” cells (**Figure 3D**, right “T”). Given that V1 is generally biased toward contralateral inputs, this may be interpreted that the more complex texture stimuli require more V2 processing.

We calculated a binocular preference index by comparing binocular responses to the sum of the ipsi and contra monocular responses. Overall, a large majority of cells in all classes, except inhibited, predominately had binocular responses, that is their responses was greater than the preferred eye (**Figure 3E, F;** more than 50%, congruent binocular preference). However, only the orthodromic cell class contained a subpopulation of cells the exhibited “strongly binocular” responses, that is, their binocular response was great than the sum of the monocular responses (**Figure 3F;** more than 100%). This is consistent with previous findings in macaque suggesting that V1->V2 projecting cells are more strongly binocular than those that receive feedback from V2[26].

Of note, the V2 location we recorded from had a disproportionately high number of ipsilateral preferring cells while the V1 area we stimulated predominately consisted of contralateral cells. The tree shrew visual cortex includes a representation of ipsilateral visual field that is highly compressed relative to that of the contralateral[27,28]. Our data using microstimulation to map V1-

V2 connections therefore suggests that the contralateral V1 sends and receives connections with the ipsilateral V2. Since we observed in V2 a strong ipsilateral bias, this suggests that in addition to our observed V1 contralateral inputs these cells must converge additional sources of inputs about the ipsilateral eye.

A potential hypothesis for the function of feedback projections is that they may provide information to upstream regions under conditions of conflicting signals[29]. Thus, binocular responsive neurons in V2 that converges information from both eyes may function in resolving conflict when different types of stimuli are presented to each eye. Using the anaglyph googles, we next presented conflicting stimuli with different visual stimuli to each eye: either static gratings or textures. We compared the responses under conditions with conflict binocular presentations to that of coherent binocular presentations. No strong difference was observed for gratings under conflict **(Figure 3E, F** offset from the diagonal). Surprisingly, many cells across all classes, except inhibited, were suppressed by conflict for texture **(Figure 3G).**

DISCUSSION

Our data demonstrate that V2 neurons in the tree shrew visual cortex can be classified into distinct anatomical and functional subtypes based on their responses to V1 stimulation. These findings highlight the complexity of visual information processing across feedforward and feedback pathways, shedding light on the interplay between local computations in V2 and their influence on V1[26]. The structure and organization of the cortical hierarchy must be understood within the framework of feedforward and feedback circuitry in the visual system[13]. At each stage, visual processing involves increasing abstraction and integration of information. Theories about feedback connections propose that feedback neurons might carry unique, feature-specific information computed in downstream areas to refine feedforward processing[11,14]. Adding to this body of work, our results show that V2 feedback neurons exhibit functional properties that align closely with other V2 neurons across a battery of visual stimuli. This suggests that the information conveyed back to V1 represents the diversity of visual information processed in V2 rather than a subset of specific features. In this sense, V2 feedback neurons may contribute to maintaining a holistic representation of V2 activity within V1 rather than transmitting specialized signals.

Despite their functional similarities, V2 feedback neurons exhibited distinct spatial properties compared to feedforward neurons. The receptive field properties of a neuron arise from the type, number and location of its inputs[30]. This is apparent along the visual hierarchy as small receptive fields tuned to simpler stimuli are combined along successive stages to form larger receptive fields tuned to more complex stimuli[31]. However, our findings indicate that V2 feedback neurons have smaller receptive fields than their V2 counterparts and rather more closely match their V1 inputs. This suggests that they contain higher-resolution information and can exert a spatially restricted influence on V1. Our data agree with the numerosity of these feedback projections as they densely tile visual space and visual features. We also found that the location of this feedback receptive field is offset relative to its V1 feedforward input, likely corresponding to the surround as has been shown in previous studies[20]. Through this organization, V2 excitatory neurons can suppress or potentially facilitate the responses of V1 neurons[32].

The visual cortex contains binocular fields that function to construct three-dimensional representations by aligning the features from the two eyes and computing visual depth through stereopsis[33]. The matching process first emerges in V1 and is further elaborated in V2[26,34,35]. However, presenting different stimuli to each eye prevents this matching and results in an alternating visual perception between the two competing stimuli known as binocular rivalry[36].

When presented with conflicting texture stimuli across the two eyes, tree shrew V2 neurons, including feedback cells, were suppressed. The suppressive effect may be computed at the level of V2 or could arise in V2 from top-down processing. This V2 conflict-induced suppression can directly influence V1 processing via V2 feedback neurons, possibly contributing to error correction or top-down resolution of visual conflict.

Notably, V2 cells were predominately ipsilaterally dominant (i.e. higher responses to ipsilateral visual stimuli than contralateral). In contrast, the connected V1 stimulation site was contralaterally dominant. This suggests that visual information in these V2 neurons is not solely inherited from these V1 inputs but is likely also derived from other cortical or subcortical sources. Our findings suggest a unique role for V2 neurons to modulate the opposite eye representation in cases with contradictory visual inputs. Whether this feedback contributes to the suppression of conflicting binocular representations as has been observed in cases of strabismus[37,38] requires further investigation. While our findings reveal new insights into the functional diversity of V2 neurons and their interactions with V1, they also raise questions about the specific mechanisms by which feedback influences are integrated into the broader visual processing hierarchy. Future studies should investigate the role of V2 feedback in the resolution of perceptual conflict, exploring how these neurons contribute to the coherence and stability of visual experience.

METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experimental procedures were approved by the Caltech Institutional Animal Care and Use Committee and conformed to local and US National Institutes of Health guidelines, including the US National Institutes of Health Guide for Care and Use of Laboratory Animals. Tree shrews (*Tupaia Belangeri*) used in this study (n=3), both male and female, were 6 months to 2.5 years old and weighted between 150 to 300 g. Animals were singly housed in a 12-hour light/dark cycle in the animal room. Their food and water aliquots were given *ad libitum*.

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

EXPERIMENT

Surgeries

Tree shrews were injected with a preoperative dose of dexamethasone (5 mg/kg, subcutaneously (s.c.)) and mannitol (1 mg/kg, s.c.) to reduce swelling. Animals were anesthetized with a cocktail of fentanyl, midazolam, and dexdomitor (FMD, fentanyl 0.05 mg/kg, midazolam 5.0 mg/kg, dexdomitor 0.25 mg/kg, s.c.), shaved, and positioned into a stereotaxic frame. Topical lidocaine gel (2%) was applied on the head and ears to prevent discomfort from ear-bars and eye lubricant was used to maintain hydration and clarity of eyes during surgical procedures. Levels of anesthesia, breathing, SpO2, and heart rate were monitored throughout the entire procedure and body temperature was maintained with a heating pad at 37.5° C. An incision on the scalp was performed and both skin and muscles were retracted. The exposed skull was levelled using the stereotaxic device with respect to bregma and lambda (pitch, roll, and yaw). After alignment, locations of the craniotomies for electrophysiological recordings were marked on the skull and a custom stainless steel headplate was secured to the skull using clear C&B Metabond (Parkell). A layer of Kwik-Cast (World Precision Instruments) was added on top the skull and a 3D printed custom cap was secured to the headplate to protect the brain and keep debris out. The anesthesia was reversed with an injection of atipamezole-flumazenil (atipamezole 1.25 mg/kg, flumazenil 0.25 mg/kg, s.c.) and the animal was recovered for at least 3 days before following procedures and recordings. One day before electrophysiological recordings in a new brain location, tree shrews were once again anesthetized and monitored as described above. Using the marked locations on the skull, small (up to 1.5mm of diameter) craniotomies were drilled and durotomy was performed. Through a small hole situated anterior of bregma, a 32 AWG chlorinated silver wire (A-M system) with a pre-soldered gold pin was implanted just above the brain surface and cemented to the skull to provide chronic grounding. A drop of silicone oil (30,000 cSt, Aldrich) was added over the holes to prevent the brain from drying, a new layer of Kwik-Cast was applied on top of it, and the 3D printed custom cap secured to the headplate. Anesthesia was reversed and the animal was recovered as previously described.

Electrophysiological recordings

All electrophysiological recordings were made using high channel-count, silicon, “Neuropixels 1.0” probes configured to always acquire from the first 384 electrodes closest to the tip, providing a 3.84 mm of tissue coverage. The reference and the ground contacts on the Neuropixels probes were permanently soldered together. Recordings were made using an external reference configuration achieved by connecting the probe reference to the chronically implanted silver wire on the skull which conductivity was routinely checked before recording with a multimeter. Each Neuropixels was mounted on a 3-axis micromanipulator (New Scale Technologies) that was in turn mounted on the underside of a semicircular platform, allowing simultaneous insertion of up

to four probes at different angles. Before the first insertion of a probe in a new location, DiI (1 mM in ethanol) was used to coat the shank, allowing subsequent probe track localization during ex vivo imaging. Neural signals were acquired at 30 kHz using Open Ephys software [39]. After the tip of each probe touched the surface of the brain, they were lowered to target at an average speed of 100 µm/min to avoid damage and let them settle for 15 minutes after reaching the target depth. Cameras were used to monitor animals during experiments and to ensure a continuative viewing of the visual stimuli presented during neural signals acquisition. After each recording experiment, probes were slowly retracted and immersed in 1% Tergazyme solution to remove tissue and silicone oil residues.

Histology

After electrophysiological recordings or tracer expression, histological verification was performed for all tree shrews. Tree shrews were given Ketamine + Xylazine and perfused transcardially with 0.9% saline, followed by 4% paraformaldehyde (PFA) in 1X PBS. Brains were extracted and post-fixed overnight in 4% PFA at 4°C. The brains were then transferred to 30% sucrose for cryoprotection and sectioned coronally at 100 µm on a cryostat (Leica Biosystems). Sections were washed with 1xPBS and then incubated for 30 minutes at room temperature in in DAPI/PBS (0.5 µg/ml) for counterstaining. Sections were then mounted on slides and imaged with an epifluorescence microscope (Olympus VS120).

VISUAL STIMULATION

Visual stimuli presentation

Visual stimuli were generated and presented using custom Python scripts. Head-fixed tree shrews passively viewed a battery of visual stimuli displayed using a ViewSonic monitor (70x39 cm, 60 Hz refresh rate, 1,920 × 1,080 pixels). The monitor was centered in front of the animals at 25 cm distance. Stimuli were presented at 3 Hz, 167 ms of image presentation interleaved with 167 ms of a grey screen. Three classes of visual stimuli were used in each experiment: static gratings, naturalistic textures and noise, and 1593 objects. In addition, “local sparse noise” stimuli were used to map neurons’ receptive fields.

Local Sparse Noise

The screen was divided into a grid of 4x3 squares. In consecutive frames (100 ms), sparse white or black dots (5 degrees square) were presented, one dot in each grid square. The locations of the dots within each rectangle were pseudo-randomly distributed to avoid spurious correlation between distant parts of the visual field[40]. To avoid interference between reconstruction of “On” and “Off” RFs, each presented stimulus frame comprised either all black or all white dots on a grey field. A reduced version of this stimulus (with fewer frames) was used at the beginning of each experiment and analyzed immediately to allow placement of “faces and objects” stimuli in the centroid of the receptive fields for that recording session.

Static Gratings

We presented full field sinusoidal gratings, varying in orientation (6 evenly spread angles), spatial frequency (5 values between 0.1 and 1.6 cycles/degree), and phase (4 positions), for a total of 120 different stimulus conditions. Each image was presented 5 times.

Naturalistic textures

We presented images from two subclasses: naturalistic textures and a control set comprising spectrally matched noise. The naturalistic textures images were organized as 15 families of 5 similar images. Texture images reproduced the statistical dependencies found in natural texture scenes[1]. Each of the 150 images in the stimulus set was presented 5 times. We used two types

of visual stimuli similar to ones previously used in primate studies: one set consisted of 15 families of texture images each comprising 5 closely related image samples of the same texture. These images reproduced statistical dependencies found in natural texture scenes[41,42]. A control set consisted of noise images spectrally matched to each of the texture families.

ELECTRICAL STIMULATION

Electrical stimulation experiment

A bipolar platinum/iridium electrode (100 kΩ, blunted tip, Microprobes for Life Science, Gaithersburg, MD) was inserted into V1 to electrically stimulate both V1 neurons orthodromically and V2 neurons antidromically. Neural activity was simultaneously recorded in V2 with a Neuropixels probe. To electrically excite regions around the electrode, biphasic pulses (200 μs duration per phase, 50 to 450 μA) were delivered using a custom-built isolated pulse stimulator (manuscript in preparation).

Artifact removal

The electrical artifact arising during electrical stimulation and simultaneous recording of signals from a silicon probe was removed offline using the SALPA algorithm[16]. Together with a careful choice of stimulation waveforms, this nonlinear digital filter enabled the detection of spikes as early as 1 ms after stimulation.

DATA ANALYSIS

Preprocessing and Spike sorting

Neural signals from electrophysiological recordings were preprocessed by subtracting the median calculated within each group of 24 channels from the data to eliminate common-mode noise. The median subtracted data was sent to Kilosort2 which in addition to the group median subtraction applied a high-pass filter (150-Hz), followed by whitening in blocks of 32 channels. All spike data were sorted using the off-line spike sorting algorithm Kilosort2[43]. The cluster automatically labelled by Kilosort algorithm as “good” were in turn manually curated by hand and further analyzed with Phy2.

Visually responsive cells

A cell was deemed responsive to a particular class of stimuli (either gratings, textures, and noise, or faces and objects) if its average firing rate in the 100 ms following stimuli of that class exceeded the expectation value based on a Poisson model trained on the firing rate in the 50 ms before all the stimuli of that class. To be included in the “responsive fraction” in figures 3b, 4b, and 5b, a cell’s average response had to exceed the baseline by at least 5 standard deviations. For the “faces and objects,” the total time elapsed between the first and the last of the 10 blocks of visual presentations was so long that stability of responses was a concern. Accordingly, we additionally preprocessed these data to analyze only those blocks in which the responses were stable for a given cell. For each block, we extracted the average waveform of all the spikes from the given cell and calculated its peak-to-peak amplitude. We then picked the third largest amplitude among the blocks and set an amplitude threshold at 0.6x this value. We counted for each block the number of individual spikes with amplitudes exceeding this threshold. We calculated the mean and standard deviation of these counts among blocks, and excluded from analysis any block in which the count was over two standard deviations below the mean. In all cases except Figure 1e, results are expressed as a percentage of visually responsive cells, i.e., of cells that respond to any of the stimulus classes.

Receptive field analysis

The receptive field size, amplitude, and quality was obtained by first calculating a 2D histogram of spike counts at each of 576 locations on the monitor (32 x 18 matrix). We modeled these histograms as a 2D Gaussian peak on top of a constant baseline. To prevent overfitting, the shape of the Gaussian was forced to be circular rather than elliptic. A cell was considered to possess an ("ON" or "OFF") receptive field if the number of spikes within the Gaussian peak exceeded expectation from a null model. Specifically, we calculated the expected number of spikes that would be elicited by ("ON" or "OFF") stimuli within a 10-degree radius from the center of the Gaussian under the null model of the baseline as well as the actual number of spikes elicited by stimuli within that same area. The number of standard deviations by which the actual number of spikes exceeded the null expectation was considered the "quality" of the RF. Only cells with RF quality greater than 5 were considered to possess an RF.

References

1. Shen, S. *et al.* Distinct organization of two cortico-cortical feedback pathways. *Nat. Commun.* **13**, 6389 (2022).
2. Zhang, S. *et al.* Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. *Science* **345**, 660–665 (2014).
3. Pafundo, D. E., Nicholas, M. A., Zhang, R. & Kuhlman, S. J. Top-down-mediated facilitation in the visual cortex is gated by subcortical neuromodulation. *J. Neurosci.* **36**, 2904–2914 (2016).
4. Gilbert, C. D. & Li, W. Top-down influences on visual processing. *Nat. Rev. Neurosci.* **14**, 350–363 (2013).
5. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
6. Sincich, L. C. & Horton, J. C. The circuitry of V1 and V2: integration of color, form, and motion. *Annu. Rev. Neurosci.* **28**, 303–326 (2005).
7. Wong, P. & Kaas, J. H. Architectonic subdivisions of neocortex in the tree shrew (Tupaia belangeri). *Anat. Rec.* **292**, 994–1027 (2009).
8. Sedigh-Sarvestani, M. *et al.* A sinusoidal transformation of the visual field is the basis for periodic maps in area V2. *Neuron* **109**, 4068-4079.e6 (2021).
9. Petry, H. M. & Bickford, M. E. The Second Visual System of The Tree Shrew. *Journal of Comparative Neurology* vol. 527 679–693 Preprint at https://doi.org/10.1002/cne.24413 (2019).
10. Humphrey, A. L., Albano, J. E. & Norton, T. T. Organization of ocular dominance in tree shrew striate cortex. *Brain Res.* **134**, 225–236 (1977).
11. Briggs, F. Role of Feedback Connections in Central Visual Processing. *Annual Review of Vision Science* **6**, 1–22 (2020).

12. Kennedy, H. & Bullier, J. A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. *J. Neurosci.* **5**, 2815–2830 (1985).
13. Van Essen, D. C. & Maunsell, J. H. R. Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci.* **6**, 370–375 (1983).
14. Gilbert, C. D. & Sigman, M. Brain states: top-down influences in sensory processing. *Neuron* **54**, 677–696 (2007).
15. Jun, J. J. *et al.* Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
16. Wagenaar, D. A. & Potter, S. M. Real-time multi-channel stimulus artifact suppression by local curve fitting. *J. Neurosci. Methods* **120**, 113–120 (2002).
17. Wagenaar, D. A., Nadasdy, Z. & Potter, S. M. Persistent dynamic attractors in activity patterns of cultured neuronal networks. *Phys. Rev. E* **73**, 1–8 (2006).
18. Rolston, J. D., Wagenaar, D. A. & Potter, S. M. Precisely timed spatiotemporal patterns of neural activity in dissociated cortical cultures. *Neuroscience* **148**, 294–303 (2007).
19. Bishop, P. O., Burke, W. & Davis, R. Single-unit recording from antidromically activated optic radiation neurones. *J. Physiol.* **162**, 432–450 (1962).
20. Keller, A. J., Roth, M. M. & Scanziani, M. Feedback generates a second receptive field in neurons of the visual cortex. *Nature* **582**, 545–549 (2020).
21. Fişek, M. *et al.* Cortico-cortical feedback engages active dendrites in visual cortex. *Nature* **617**, 769–776 (2023).
22. Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**, 574–591 (1959).
23. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
24. Zeki, S. M. Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *J. Physiol.* **277**, 273–290 (1978).

25. Muly, E. C. & Fitzpatrick, D. The morphological basis for binocular and ON/OFF convergence in tree shrew striate cortex. *J. Neurosci.* **12**, 1319–1334 (1992).
26. El-Shamayleh, Y., Kumbhani, R. D., Dhruv, N. T. & Movshon, J. A. Visual response properties of V1 neurons projecting to V2 in macaque. *Journal of Neuroscience* **33**, 16594–16605 (2013).
27. Bosking, W. H., Beauchamp, M. S. & Yoshor, D. Electrical Stimulation of Visual Cortex: Relevance for the Development of Visual Cortical Prosthetics. *Annual Review of Vision Science* **3**, 141–166 (2017).
28. Bosking, W. H., Kretz, R., Pucak, M. L. & Fitzpatrick, D. Functional specificity of callosal connections in tree shrew striate cortex. *J. Neurosci.* **20**, 2346–2359 (2000).
29. Bastos, A. M. *et al.* Canonical Microcircuits for Predictive Coding. *Neuron* **76**, 695–711 (2012).
30. Seabrook, T. A., Burbridge, T. J., Crair, M. C. & Huberman, A. D. Architecture, Function, and Assembly of the Mouse Visual System. *Annu. Rev. Neurosci.* **40**, 499–538 (2017).
31. Felleman, D. J. & Van Essen, D. C. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb. Cortex* (1991).
32. Schwabe, L., Obermayer, K., Angelucci, A. & Bressloff, P. C. The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. *J. Neurosci.* **26**, 9117–9129 (2006).
33. Read, J. C. A. Binocular vision and Stereopsis across the animal kingdom. *Annu. Rev. Vis. Sci.* **7**, 389–415 (2021).
34. Thomas, O. M., Cumming, B. G. & Parker, A. J. A specialization for relative disparity in V2. *Nat. Neurosci.* **5**, 472–478 (2002).
35. Barlow, H. B., Blakemore, C. & Pettigrew, J. D. The neural mechanism of binocular depth discrimination. *J. Physiol.* **193**, 327–342 (1967).

36. Hesse, J. K. & Tsao, D. Y. A new no-report paradigm reveals that face cells encode both consciously perceived and suppressed stimuli. *Elife* **9**, e58360 (2020).
37. Frisby, J. P., Mein, J., Saye, A. & Stanworth, A. Use of random-dot sterograms in the clinical assessment of strabismic patients. *Br. J. Ophthalmol.* **59**, 545–552 (1975).
38. Economides, J. R., Adams, D. L. & Horton, J. C. Interocular suppression in primary visual cortex in strabismus. *J. Neurosci.* **41**, 5522–5533 (2021).
39. Siegle, J. H. *et al.* Open Ephys: An open-source, plugin-based platform for multichannel electrophysiology. *J. Neural Eng.* **14**, (2017).
40. Siegle, J. H. *et al.* Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92 (2021).
41. Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P. & Movshon, J. A. A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.* **16**, 974–981 (2013).
42. Portilla, J. & Simoncelli, E. P. *A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients*. vol. 40 49–71 (2000).
43. Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris, K. Fast and accurate spike sorting of high-channel count probes with KiloSort. *Adv. Neural Inf. Process. Syst.* **29**, 4448–4456 (2016).

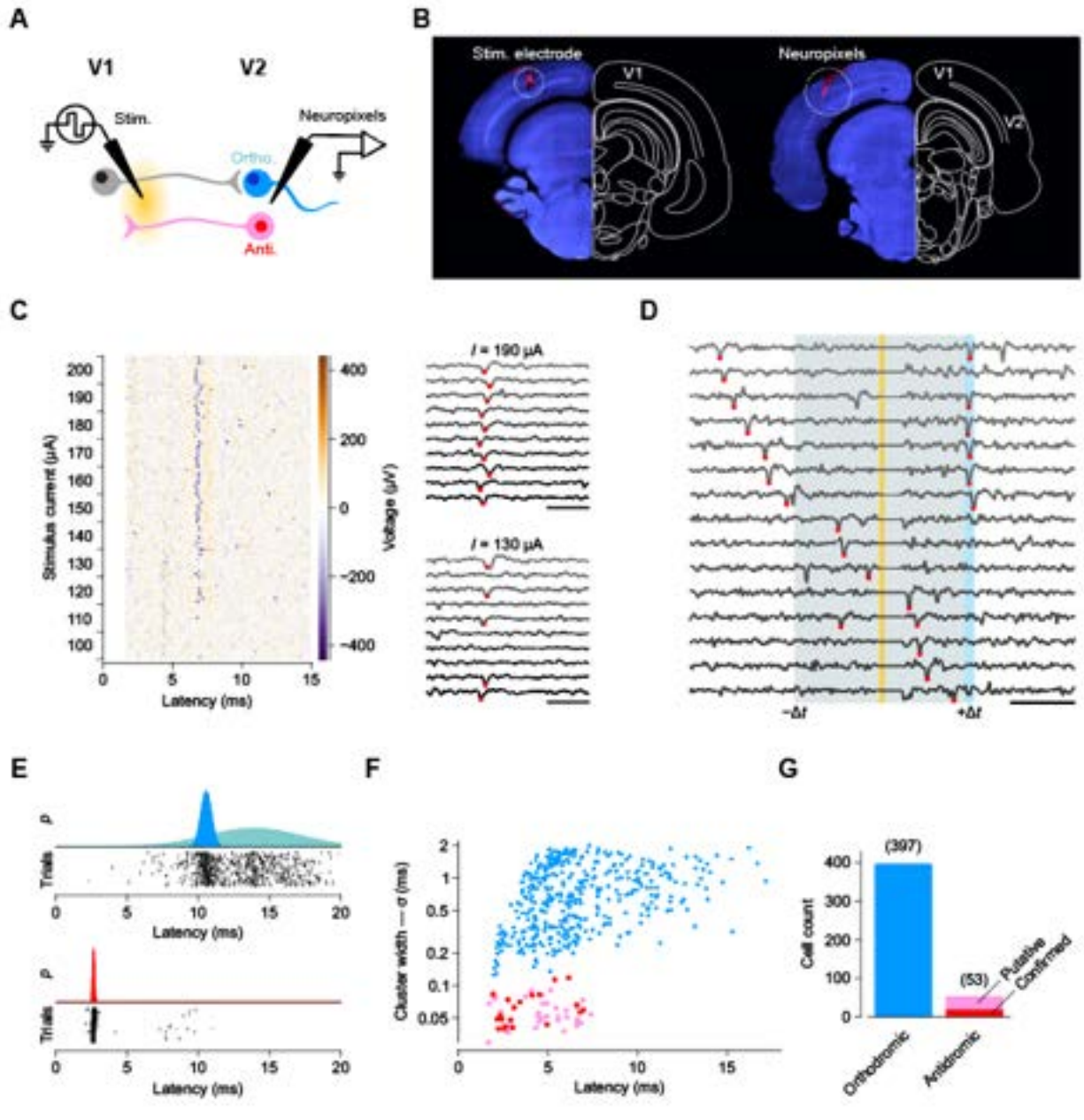

A
V1
V2
Stim.
Neuropixels
Ortho.
Anti.
B
Stim. electrode
V1
Neuropixels
V1
V2
C
Stimulus current (μA)
Latency (ms)
Voltage (μV)
I = 190 μA
I = 130 μA
D
−Δt
+Δt
E
Trials
Latency (ms)
F
Cluster width — σ (ms)
Latency (ms)
G
Cell count
(397)
(53)
Putative
Confirmed
Orthodromic
Antidromic

Figure 1: Combining Neuropixels recording with bipolar electrode stimulation enables identifying feedforward and feedback neurons.

(A) Schematic of circuit under study in the tree shrew visual cortex. Electrical stimulation with Pl/Ir bipolar electrodes in primary visual cortex V1. Recording with Neuropixels probe in secondary visual area V2. The blue V2 cell receives feedforward signals from V1; the pink V2 cell sends feedback signals to V1.

(B) Coronal section illustrating placement of stimulating and recording electrodes in V1 and V2 respectively, marked with DiI (*red*). Scale bar: 1 mm.

(C) Ramp test for an example V2 neuron. Raster plot of responses of a representative cell to biphasic pulses of increasing current amplitude. High-current pulses (above the threshold necessary to evoke an action potential) activate the cell 100% of the time (*top right inset*). Near-threshold pulses activate the cell unreliably (i.e., only in a fraction of trials; *bottom right inset*). Scale bars: 2 ms.

(D) Collision test for an example V2 neuron. Electrically evoked spikes occurred, on average, at $+\Delta t$ (blue bar) relative to the electrical pulse (yellow bar). A subset of trials is plotted in which the cell fired spontaneously around the time of the pulse. (Trials are sorted in order of time of occurrence of spontaneous spikes.) When a spontaneous action potential occurs in the interval $[-\Delta t, +\Delta t]$ (shaded region) it collides with the electrically evoked action potential in the axon, which then does not reach the soma, resulting in the absence of a recorded spike at $+\Delta t$.

(E) Example of a V2 orthodromic cell (top) and a V2 antidromic cell (bottom), classified using FSMEM [an algorithm that fits a mixture of gaussian onto the responses of a neuron to electrical stimulation] (see STAR Methods). Dots: recorded spikes; shading: fitted mixture of gaussians.

(F) Scatter plot of the latency and width of all peaks obtained using FSMEM. Blue: putative orthodromic neurons. Red: antidromic cells with confirmed collision test. Pink: putative antidromic cells with spontaneous firing rate too low for applying the collision test.

(G) Number of orthodromic and antidromic cell identified across 10 recordings.

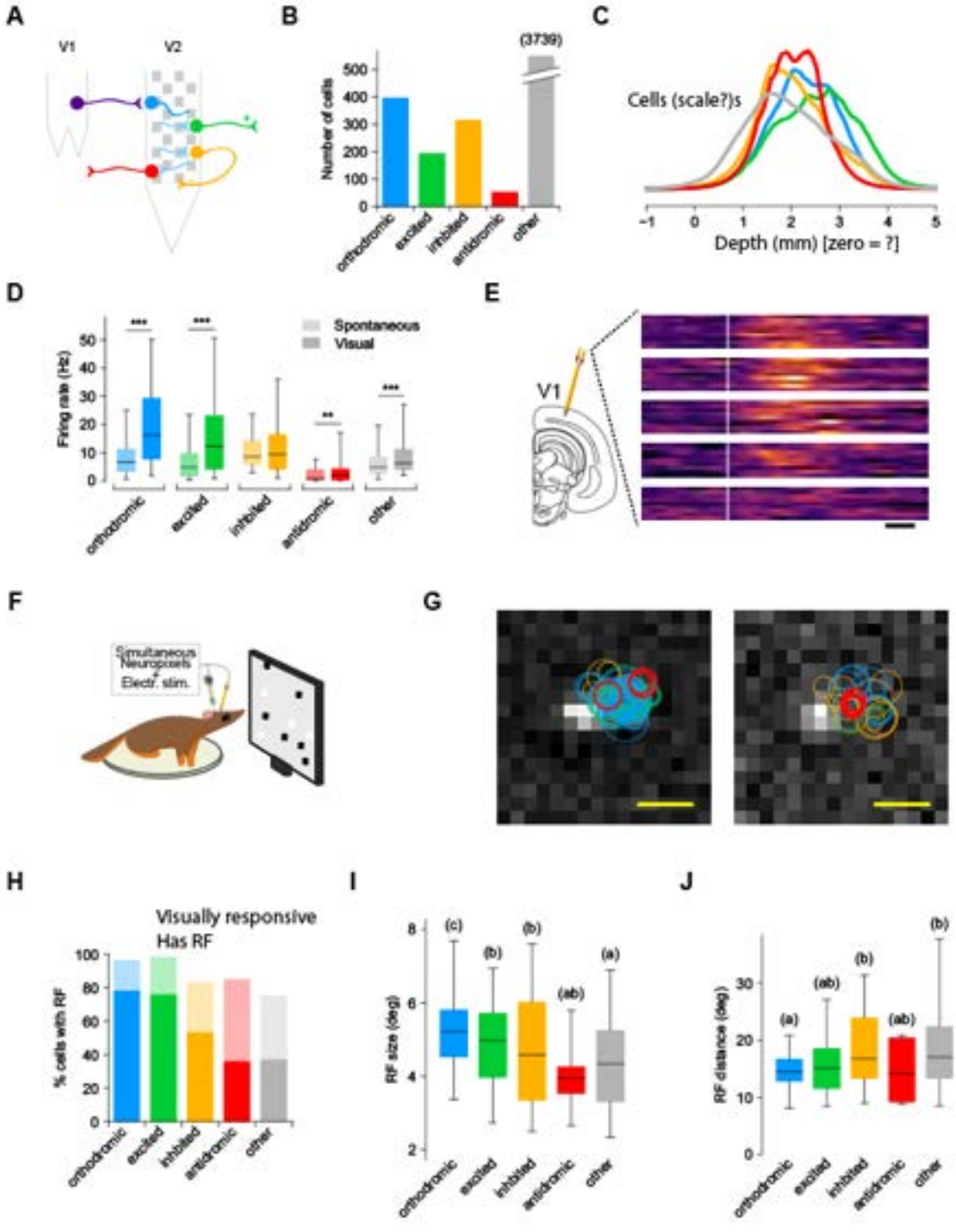

A
V1
V2
B
(3739)
Number of cells
orthodromic
excited
inhibited
antidromic
other
C
Cells (scale?)s
Depth (mm) [zero = ?]
D
Firing rate (Hz)
Spontaneous
Visual
E
V1
F
Simultaneous Neuropixels + Electr. stim.
G
H
Visually responsive
Has RF
% cells with RF
I
RF size (deg)
J
RF distance (deg)

Figure 2: Feedback cells have the smallest receptive fields of all classes of V2 cells.
(A) Recorded cells can be classified into several categories based on their responses to electrical stimulation: *Orthodromic* (blue) and *antidromic* (red) as in Fig. 1; *excited* (green) are non-orthodromic cells in which neuronal activity increased after electrical pulses; *inhibited* (yellow) are cells in which neural activity decreased after electrical pulses.
(B) Experimental configuration: Presentation of a battery of visual stimuli to a head-fixed tree shrew with a Neuropixels probe in V2 and a Pt/Ir stimulation electrode in V1.
(C) Range of spontaneous and visually evoked firing rates for each class of cells. Boxes represents 25^{th}, 50^{th}, and 75^{th} percentiles; whiskers 5^{th} and 95^{th} (also in G and H).
(D) When not actively used for stimulation, the V1 electrode can be used to record multiunit activity from V1. Schematic illustrating electrophysiological recording from the electrode in V1 and actual instantaneous energy in V1 for gratings presentations. Scale bar: 10 ms.
(E) Example receptive field (RF) of V1 multiunit activity (grayscale map) with sizes and positions of V2 RFs overlaid. Colors represent V2 cell classes as in A and C. Scale bar: 20 degrees.
(F) Percentage of cells within each class that had a defined RF.
(G) RF size distribution within each class.
(H) Distance of the RF center for each cell from the V1 center of mass in the same experiment.

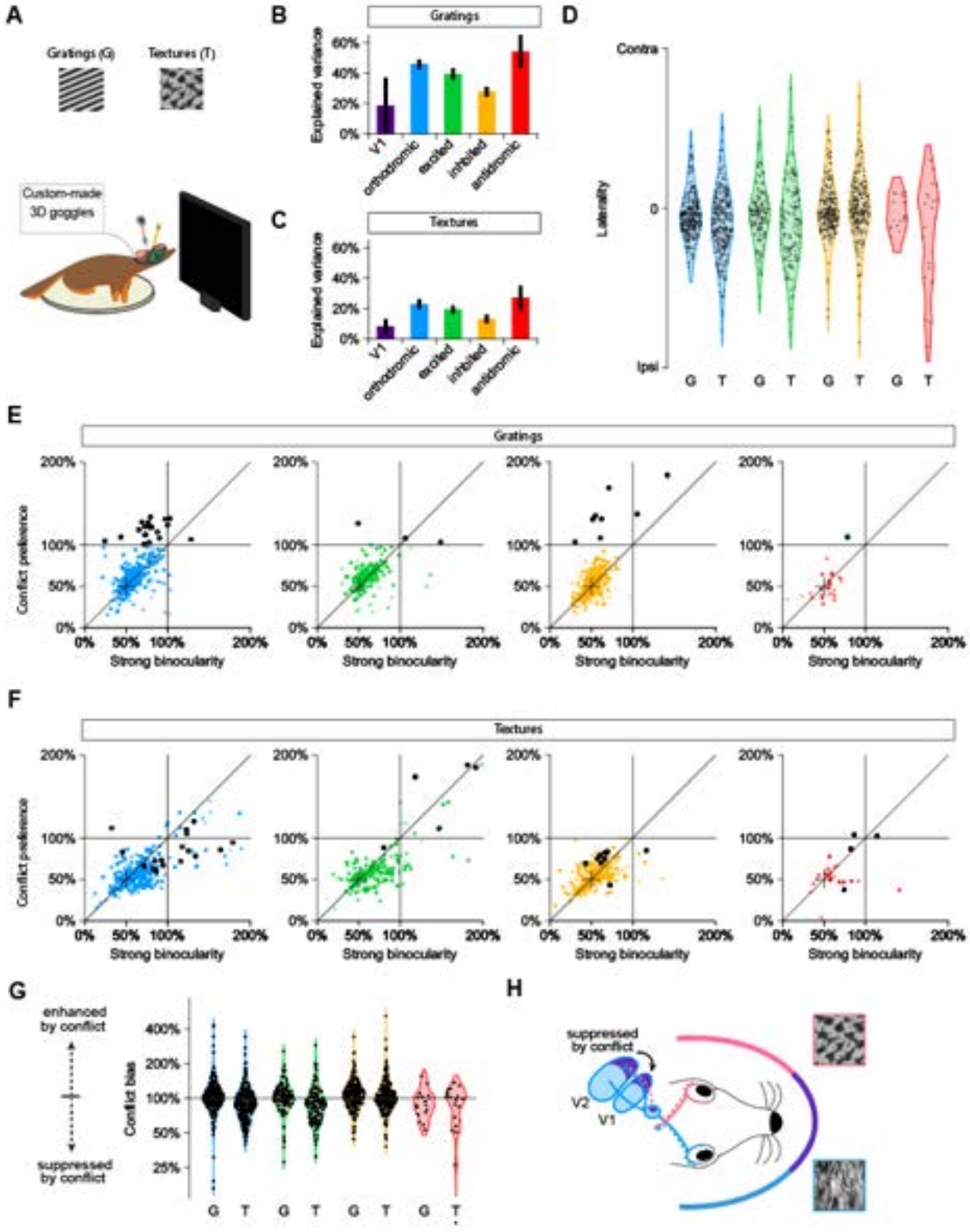

A
Gratings (G)
Textures (T)
Custom-made 3D goggles
B
Gratings
Explained variance
60%
40%
20%
0%
V1
orthodromic
excited
inhibited
antidromic
C
Textures
D
Contra
Laterality
0
Ipsi
G T G T G T G T
E
Gratings
Conflict preference
200%
100%
50%
0%
0% 50% 100% 200%
Strong binocularity
F
Textures
G
enhanced by conflict
suppressed by conflict
Conflict bias
400%
200%
100%
50%
25%
H
suppressed by conflict
V2
V1

Figure 3: Cell class defined by responses to electrical stimulation is predictive of responses to binocular conflict stimuli.
(A) Experimental configuration: Presentation of static grating or naturalistic texture stimuli to a head-fixed tree shrew with anaglyph googles while recording in V2 and stimulating in V1.
(B) Percentage of variance of individual cells' responses explained by orientation of gratings.
(C) Same, but for naturalistic texture stimuli.
(D) Laterality of individual cell's responses for stimuli presented to the contralateral versus ipsilateral eye.
(E) Responses to static gratings presented to both eye that were either matching (congruent binocular preferring) or not (conflict binocular preferring).
(F) Same, but for naturalistic texture stimuli.
(G) Relative response to conflict compared to that in congruent conditions.
(H) Summary schematic of findings.

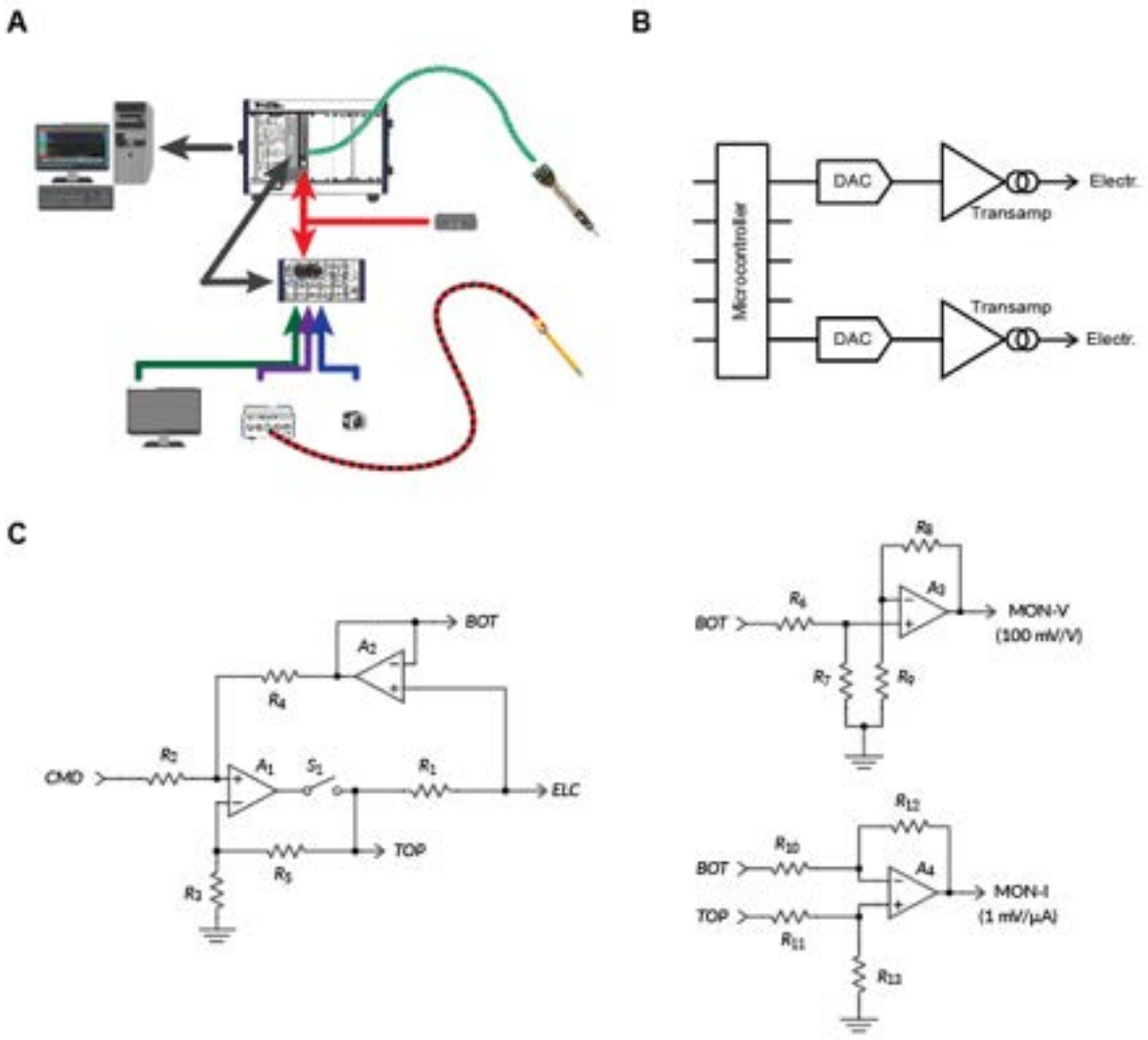

A
B
Microcontroller
DAC
Transamp
Electr.
DAC
Transamp
Electr.
C
BOT
A2
R4
CMD
R2
A1
S1
R1
ELC
TOP
R5
R3
R8
A3
R6
BOT
MON-V
(100 mV/V)
R7
R9
R12
R10
BOT
A4
MON-I
(1 mV/µA)
TOP
R11
R13

Figure S1: High current electrical microstimulator.
(A) Schematic illustrating the simultaneous electrical stimulation and Neuropixels recording setup.
(B) Simplified schematic of our stimulator: A Teensy microcontroller is used to independently set the current through a bipolar pair of Pt/Ir microelectrodes.
(C) Detail of transimpedance amplification. The voltage from the DAC is mirrored across R1, setting the current through the electrode. The instantaneous electrode voltage and current is monitored at VELC and IELC respectively. A digital switch (S1) reduces noise coupling when the stimulator is not delivering current pulses.

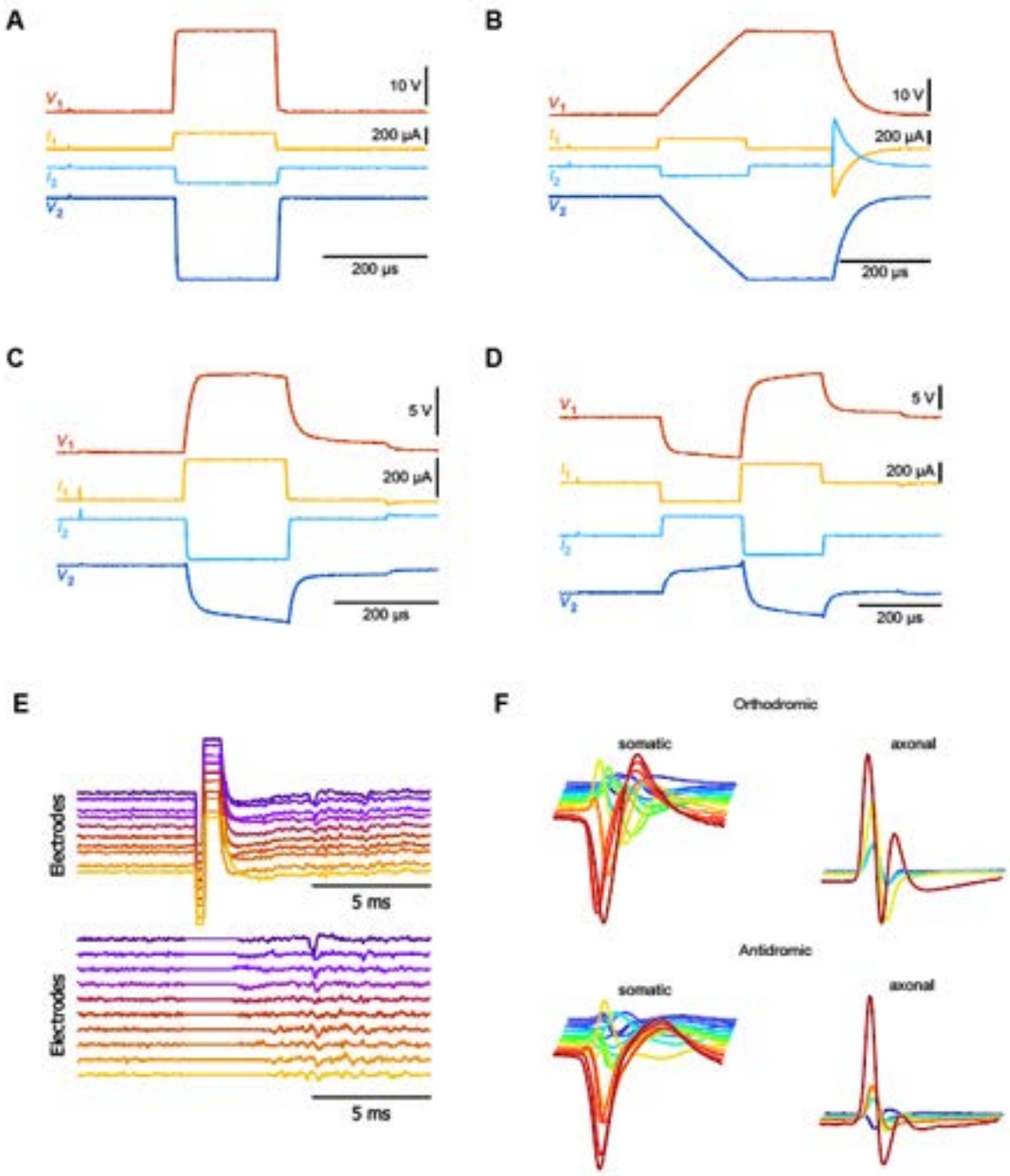

A
10 V
200 µA
200 µs
B
10 V
200 µA
200 µs
C
5 V
200 µA
200 µs
D
5 V
200 µA
200 µs
E
Electrodes
5 ms
Electrodes
5 ms
F
Orthodromic
somatic
axonal
Antidromic
somatic
axonal

Figure S2: Oscilloscope traces.
(A) Oscilloscope traces of electrode voltages and currents in each of the paired electrodes during a bipolar monophasic current pulse in saline.
(B) Same for a biphasic stimulus.
(C) Results of a current pulse delivered to a pair of 100-kOhm resistors rather than Pt/Ir electrodes.
(D) Same, but using 1-nF capacitors. (The triangular current pulse on the right is not delivered to the electrode, but results in the monitor circuit from opening switch S1.)
(E) Examples of electrode traces before (top) and after (bottom) electrical artifact suppression with SALPA.
(F) Traces for somatic (left) and axonal (right) action potentials recorded with Neuropixels probe.

Chapter IV

FUTURE DIRECTIONS

We have shown that the tree shrew carries a compressed hierarchy for form processing compared to primates, in which key functions of primate posterior IT cortex are already carried out by area V2. Future studies are needed to understand the processing of these object responsive cells in tree shrew V2 with regards to its connections with lower and higher areas in the hierarchy[1–3]. What areas does V2 send outputs to? Is V2 able to directly control downstream motor areas in the tree shrew? Furthermore, while we did not find prominent object selectivity in areas anterior to V2 in the tree shrew, this begs the question, *what is the function of these areas?* Are they involved in multi-modal processing[1]? Future studies exploring tuning in these areas in naturalistic, multi-modal environments may shed new light on these areas.

We found that combining electrical microstimulation and high-density Neuropixels probes is an effective method for identifying the directionality of projections within a surveyed area. Given the broad applicability of these tools[4], it can be applied across species in tree shrews and primates. This opens the way to directly test the role of feedforward and feedback in both species for the for construction of visual percepts. For example, theories of conscious perception posit that feedback is necessary to resolve ambiguity[5,6]. The method we have developed now allows us to address the specific representation carried by feedback neurons, to directly test this hypothesis.

Zooming out, we have identified one interesting new point in the evolution of the primate visual system. It would be very useful to have additional points. For example, what is the functional organization of the ventral stream in mouse lemurs[7,8], a prosimian species more closely related to macaques evolutionarily? Furthermore, how do the visual processing capabilities of each species support the behaviors necessary for each species' ecological niche? In this thesis, experiments were all done under passive fixation, but relating neural circuits to behavior would be a major next step. Through such a program, we can gain a much deeper understanding of the evolution of form vision.

References

1. Wong, P. & Kaas, J. H. Architectonic subdivisions of neocortex in the tree shrew (Tupaia belangeri). *Anat. Rec.* **292**, 994–1027 (2009).
2. Petry, H. M. & Bickford, M. E. The Second Visual System of The Tree Shrew. *Journal of Comparative Neurology* vol. 527 679–693 Preprint at https://doi.org/10.1002/cne.24413 (2019).
3. Emmons, L. *Tupai: A Field Study of Bornean Treeshrews*. (University of California Press, 2000).
4. Cohen, M. R. & Newsome, W. T. What electrical microstimulation has revealed about the neural basis of cognition. *Curr. Opin. Neurobiol.* **14**, 169–177 (2004).
5. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
6. Friston, K. The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* **13**, 293–301 (2009).
7. Luongo, F. J. *et al.* Mice and primates use distinct strategies for visual segmentation. *Elife* **12**, (2023).
8. Ho, C. L. A. *et al.* Orientation preference maps in Microcebus murinus reveal size-invariant design principles in primate visual cortex. *Curr. Biol.* **31**, 733-741.e7 (2021).