Neural construction of conscious perception

Thesis by

Janis Karan Hesse

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2020
(Defended May 28th, 2020)

© 2020

Janis Hesse
ORCID: 0000-0003-0405-8632

ACKNOWLEDGEMENTS

I'd like to thank Doris Tsao for being an incredible advisor and undepletable source of ideas, advice and inspiration, for sharing her passion for science with me, and giving me the courage to ask big questions.

I am very happy about the choice of my thesis committee. Markus Meister, Ueli Rutishauser, and Ralph Adolphs contributed substantially by providing useful ideas, discussions, and criticisms throughout this thesis.

I want to thank current and past members of the Tsao lab, including Varun Wadia, Nicole Schweers, Audo Flores, Pinglei Bao, Liang She, Steven Le Chang, Xueqi Cheng Shay Ohayon, Tomo Sato, Joseph Wekselblatt, Francisco Luongo, Lu Liu, Anne Martin, Jessa Alexander, Erin Koch, Jialiang Lu, Yuelin Shi, Alex Farhang, Irene Caprara, Frank Lanfranchi, Lindsay Salay, Hongsun Guo, Abriana Sustaita, and Sebastian Moeller, who have all been very willing to offer me help whenever I needed, taught me the different techniques in the lab, and gave me great comments, ideas and discussions. I would also like to note that the work on human epilepsy patients described in Chapter VI is as much of Varun Wadia's work as it is mine.

I am grateful to have started my PhD with such a lovely cohort. My PhD would not have been as fun without Mason McGill, Vineet Augustine, Gabriela Tavares, and Ryan Cho.

I am very grateful for having met MACS, who strongly motivated me while writing this thesis.

I want to thank Angela Hesse for gifting me this life and raising me in it in such a loving way.

ABSTRACT

Out of a myriad of sensory stimulations, our brain constructs a unified, self-consistent reality that we consciously experience. Little is known about how or where in the brain's processing stream of physical input a conscious percept emerges into awareness. A remarkable property of conscious perception is that even though external input is often ambiguous, the perceptual interpretation of the world that our brain generates is consistent across multiple layers of representation, e.g., figure-ground segmentation and object identity. We thus set out to study how the interaction between different nodes in the brain generates and propagates new conscious percepts. Since the code of object identity is already well-understood, in particular for faces as reviewed in this thesis, we decided to get a handle on segmentation signals first. It turned out that consistent segmentation signals are hard to find, however, we found functionally defined modules in the brain that contained consistent cells from which figure-ground signals can be decoded. We next investigated whether face cells in object recognition areas actually encode the conscious percept of a face or are just passive filters of visual input. To distill conscious perception from other cognitive processes, such as decision making, introspection, and reporting of the percept, which often accompany new conscious percepts, we developed a no-report binocular rivalry paradigm that relies on an active fixation task rather than report, and therefore eliminates these confounding factors. We found that face patches in inferotemporal cortex indeed encode the conscious percept of a face. Using novel high-yield electrodes, we were able to decode what the animal was consciously perceiving at a given time. Preliminary and future experiments of population recordings from multiple nodes of the cortical hierarchy simultaneously promise to go beyond correlates of consciousness and reveal the mechanisms of how and where conscious percepts are constructed.

PUBLISHED CONTENT AND CONTRIBUTIONS

Hesse, Janis K., and Tsao, Doris Y. (2016). "Consistency of border-ownership cells across artificial stimuli, natural stimuli, and stimuli with ambiguous contours." In: *Journal of Neuroscience* 36.44 (2016): 11338-11349. DOI:10.1523/JNEUROSCI.1857-16.2016

J.K.H. conducted the experiments and analyzed the data. J.K.H. and D.Y.T. designed the experiments, interpreted the data, and wrote the paper.

Hesse, Janis K. and Tsao, Doris Y. (2020). " Representation of conscious percept without report in the macaque face patch network". In : *bioRxiv*. DOI: 10.1101/2020.04.22.047522. Submitted to *eLife*.

J.K.H. conducted the experiments and analyzed the data. J.K.H. and D.Y.T. designed the experiments, interpreted the data, and wrote the paper.

Hesse, Janis K. and Tsao, Doris Y. (2020). " The macaque face patch system: a turtle's underbelly for the brain". Submitted to: *Nature Reviews Neuroscience*.

J.K.H. and D.Y.T. wrote the paper.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Number *Page*

NOMENCLATURE

Binocular rivalry. The phenomenon that if two incompatible inputs are presented to the left and right eyes, respectively, rather than seeing a superimposition of the two, one’s percept usually switches between the two images.

Face patch. A cortical region in the macaque brain that responds selectively to viewed faces. Besides the six canonical face patches (PL, ML, MF, AL, AF, and AM) in each hemisphere that are found in IT cortex, face patches have been reported outside of IT, such as the upper bank of the superior temporal sulcus, perirhinal, and prefrontal cortex.

Firing rate. The number of action potentials a neuron fires per second. Action potentials or “spikes” are electrical impulses generated by neurons to send information to each other.

Functional magnetic resonance imaging (fMRI). A method to image activation of the entire brain over time. Has low temporal and spatial resolution compared to invasive electrophysiology.

Inferotemporal cortex (IT). A cortical region in the inferior convexity of the temporal lobe of the macaque brain thought to be homologous to the human ventral temporal cortex. IT is part of the ventral visual stream and important for object recognition.

Local Field potential (LFP). A signal picked up by the electrode that is generated by the current flow of several nearby neurons.

Receptive field. A portion of sensory space that can elicit sensory activation when stimulated. For visual neurons, receptive fields are usually defined as the region of visual space that can modulate activity of the neuron if a specific stimulus is presented in it.

Retinotopic cortex. A collection of visual areas that each have a retinotopic map of visual space. That is, for every position on the retina, there is a corresponding location in the visual area that represents it. Neighboring locations in the visual area correspond to neighboring areas in visual space. In general, early to mid-level regions V1, V2, V3, V3A, V4, and V4A are all considered part of retinotopic cortex, whereas the higher-level stages of the visual hierarchy such as inferotemporal cortex, where cells exhibit more spatial invariance, are not.

Chapter I

INTRODUCTION

Motivation

Having conscious experience is the most important reason why it matters to us whether we are dead or alive. Imagine having a critical car accident, and your doctor gives you two options: either you can live your normal life for another year, until the resulting brain damage takes your life; or you can be put into a dreamless coma, with little robots and artificial intelligence controlling your muscles so that you behave the same as before, but you will not have any conscious experience for the rest of your life (Revonsuo and Kamppinen, 2013). The question is a no-brainer: if you are not able to see, hear, feel, and experience the world, you may as well be dead.

Given this life-or-death importance of conscious experience, it is not surprising that humanity has been trying understand its nature, origin, or function for millennia (Parker, 1999; RielSalvatore et al., 2001). Yet, success in these domains has so far evaded humanity given that conscious experience is, while on the hand the most familiar, first-hand thing we know, at the same time also the most mysterious (Chessick, 2008). Although Aristotle claimed that to understand consciousness one must study the heart (Hicks, 2015), nowadays it is believed that, for studying consciousness, looking at the brain is not a bad idea. Therefore, in this thesis we want to describe our modest progress in understanding which signals in the brain are correlated with consciousness and give a glimpse into what mechanisms may be involved in evoking a change of conscious percept. To have a crack at the neuroscience of consciousness, one needs to decide which angle to tackle it from. Here, we investigate it through electrophysiological recordings in macaque monkeys and take their visual system as an entry point. In particular, two important visual processes, object segmentation and object recognition, will come up in this thesis.

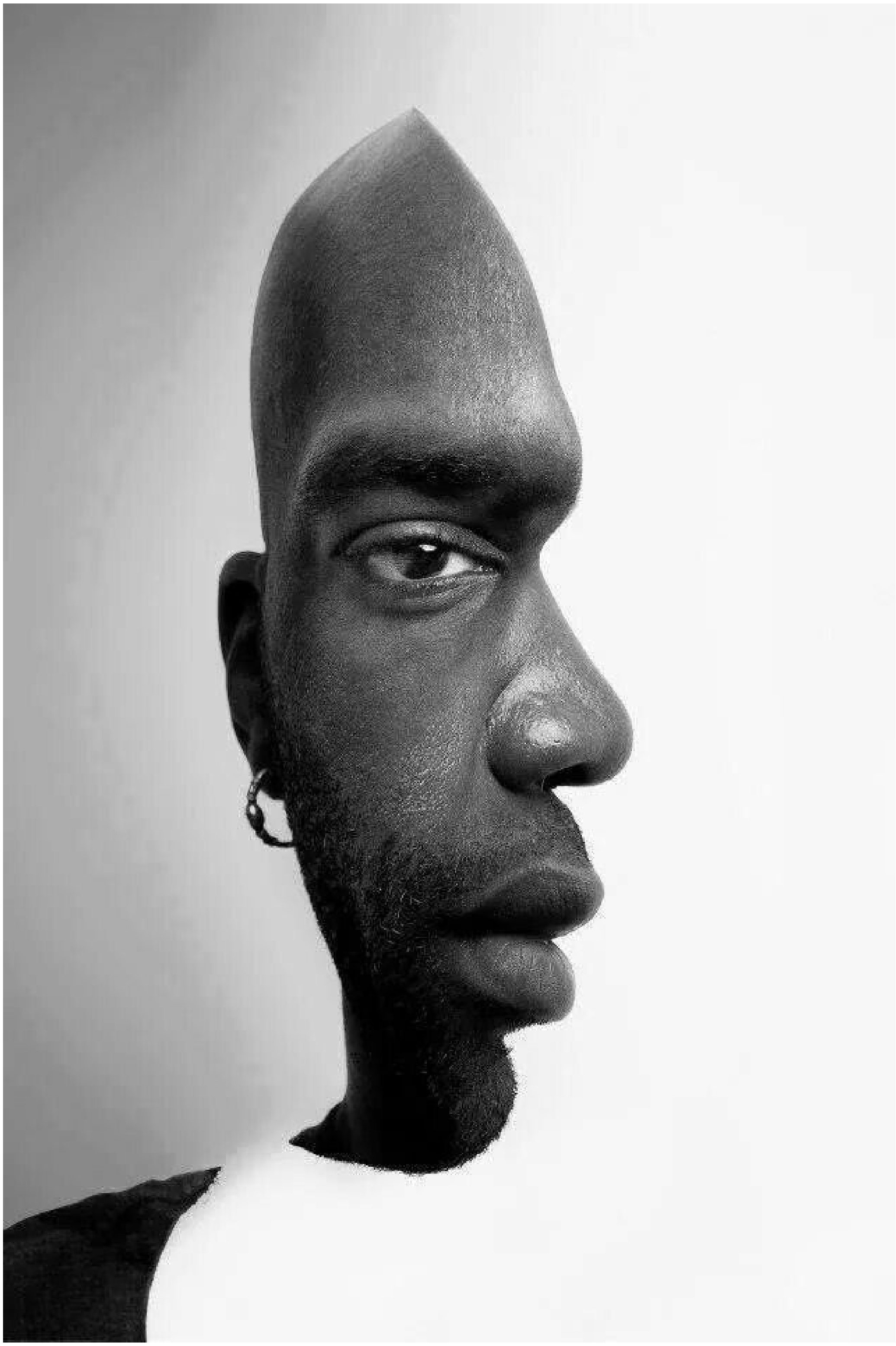


Figure I-2: *Frontal-profile illusion.*

Bistable phenomena

The saying is that a picture says more than a thousand words, which is particularly true for Fig. I-1. This was the first image my advisor and I looked at in the very beginning of my PhD; it weaves throughout like a red thread and sparked the inspiration for a fundamental question about consciousness that I am now finally close to answering at the end. When looking at the picture, you will either see a frontal face looking at you, or the profile of a face looking sideways. After looking for it for a few seconds, you percept will switch to the other interpretation, and back, and forth, in an endless dance. That is, the physical, visual input is fixed, but your conscious percept of the object is changing. This allows distilling the changes in the conscious percept from changes of physical input, which usually go hand-in-hand in the real world and are thus easily conflated. The stimulus also demonstrates another feature of conscious experience, namely its unity and self-consistency: looking at the stimulus again, you may notice that it is not just the view of the object that is changing, but with it also the figure-ground organization of the scene. If you perceive the face as facing you, you will perceive the white region on its right as background, with the border between them belonging to the outline of the face object. If you perceive the face as facing to the side, you will perceive the white region on its right it as a foreground surface occluding the face, with the border between them belonging to the outline of the occluder. Importantly, the switch of view always coincides with the switch in figure-ground organization, so that your interpretation of the scene is always self-consistent. We believe that figure-ground organization is represented in earlier, retinotopic visual cortex, whereas features of an object such as view are represented in the face patch network which is located in the higher-level inferotemporal cortex. Thus, the first question we asked during my PhD was: how do these two regions coordinate switches so that the interpretation is still consistent as a whole? What is the direction of information flow? Does the lower-level retinototopic region first switch the interpretation of figure-ground segmentation and then propagate that information to the higher-level region to switch the object view? Or does the higher-level region switch first, and then update figure-ground segmentation in lower-level regions? In broader terms, when your conscious percept changes, does that switch originate from feedforward or feedback mechanisms? At the end of this thesis, I will show some preliminary experiments that make me believe I am now closer than ever to conclusively answering this question. Note that Fig. I-1 represents only one of many examples encompassed by the general category of bistable phenomena, where the same stimulus has multiple interpretations. As an exercise, the reader can try to find the two interpretations for each of the stimuli in Fig. I-2 without reading the figure legend. For many of them, the figure-ground assignment switches as well. The next chapter outlines how, throughout this thesis, we tried to figure out the information flow across different brain areas during these perceptual switches. Indeed, understanding how brain regions interact during perceptual switches may give us general insights into how the brain arrives at an interpretation of the world that makes sense even for non-reversible stimuli. In everyday life, the world is often ambiguous, with the same input capable of having several interpretations. For example, the two processes of segmentation and recognition of objects in a visual scene constitute a bit of a chicken-egg problem and are hard to solve separately from each other. When trying to segment an ambiguous scene, it is helpful to already know the present objects and their shapes. However, for recognizing an object in clutter, one would like to consider its isolated segment alone to not get confounded by features of other objects. Thus, it is possible that corresponding visual regions recurrently interact to converge onto a self-consistent interpretation which is then broadcasted as a conscious percept into our awareness.

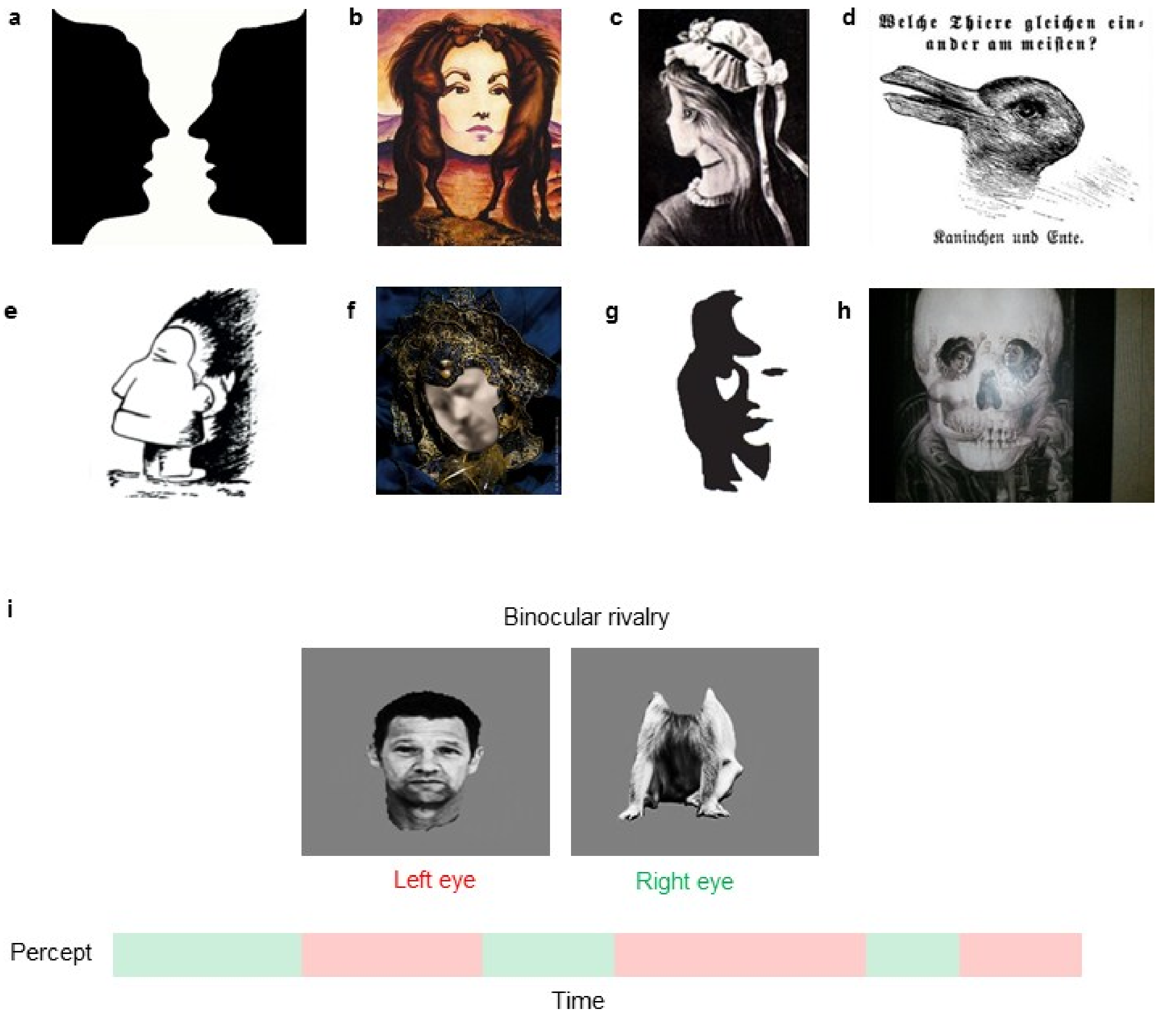


Figure I-2: Bistable phenomena. *(**a**) The famous Rubin vase switches perceptually between two black faces and a white vase (Rubin, 1980). (**b**) A female face or two horses. (**c**) An old lady or a young girl facing backwards. (**d**) A rabbit or a duck (**e**) A native American face or an Inuit facing backwards. (**f**) One face or two lovers kissing. (**g**) A woman's face or a horn player. (**h**) A skull or a woman and a man at a table. (**i**) When two incompatible stimuli such as a face and monkey body are presented to the left and right eye, respectively, one's percept is not a superimposition of the two but stochastically alternates between the two every few seconds.*

Outline

The way to answer the question of which direction information flows during switches of conscious percept seemed straightforward: if we stick an electrode into retinotopic regions that represent segmentation signals and at the same time stick an electrode into face patches, which represent object view, we should be able to find out whether during a perceptual switch segmentation areas or face patches change their activity first. As often, the actual path to the holy grail proved to be more complicated and full of detours to circumvent obstacles on the way. Through previous studies, mostly from our lab, we already had a good handle on signals in the face patches encoding face view and identity, which is reviewed in Chapter II: The Macaque Face Patch System. Thus, we decided that the first step should be to get a good handle on segmentation signals in retinotopic cortex, as described in Chapter III: Figure-Ground Segmentation. We surmised that the best bet for a reliable segmentation signal may be the so-called border-ownership cells discovered by von der Heydt and collaborators, which signal for a border in their receptive field on which side the object is that it belongs to. Contrary to our prior beliefs, consistent border-ownership cells were hard to find and not as prevalent in random locations of retinotopic regions as we thought. Yet, using fMRI we were able to find functionally defined clusters of neurons from which we could reasonably decode figure-ground segmentation. After obtaining improved segmentation signals, we wanted to know whether face patches actually do represent if the monkey was conscious of a face or not. Alternatively, face cells could just be passive filters of the visual input regardless of percept. One of the categories of switchable face stimuli that we considered using to answer this question was a type of degraded face image called Mooney faces. This endeavor is described in Chapter IV: Mooney Faces. Humans usually very rapidly recognize this stimulus as a face if it is presented upside down, concomitantly causing the perception of subjective/illusory contours of the face, however, anecdotally most monkeys do not. Hence, we asked whether we could make monkeys see the face in Mooney face stimuli and study this moment of switch in their perception. It turned out that responses to Mooney faces could be improved by animating them or adding an outline, but even after this exposure, responses to Mooney faces remained relatively low. Following this interlude, in the second half of this thesis we study the neural correlates of conscious perception more directly. In Chapter V: A Philosophical Treatise on Consciousness we distinguish between different definitions of consciousness, give a review on previous electrophysiological studies aiming to distill neural correlates of consciousness, analyze confounding factors that can foil these attempts, and dissect what it means for a signal in the brain to truly be a neural correlate of consciousness. Using these insights, we developed a no-report binocular rivalry paradigm to isolate neural correlates of consciousness as rigorously as possible and avoid confounds such as report that beset previous studies. We then employed this paradigm in monkeys and asked whether neurons in face patches actually do represent the conscious percept of a face or are just passive filters of visual input. The main results of this study, which are described in Chapter VI: Binocular Rivalry, are that face cells are indeed modulated by conscious percept even without report. Moreover, they may multiplex both conscious percept and the veridical physical stimulus, and the conscious percept can be decoded with high accuracy from a population of single neurons. In the last chapter, Chapter VII: Future Directions, we describe preliminary studies that finally hone in on answering the first question of this thesis – in which direction information flows (feedforward vs. feedback) during conscious switches – by recording with ultra-high-yield electrodes from two directly connected nodes of a hierarchy, face patches ML and AM. Moreover, we give an outlook on future experiments that could further reveal the source of conscious switches and the causal role different parts of the brain play in propagating and broadcasting a switch across the brain. These investigations hold the promise to go beyond mere correlates of consciousness and yield a mechanistic understanding of consciousness.

Chapter II

THE MACAQUE FACE PATCH SYSTEM

From: Hesse and Tsao (2020). "The macaque face patch system: a turtle’s underbelly for the brain". Submitted to: *Nature Reviews Neuroscience*.

Objects constitute the fundamental currency of our consciousness: they are the things that we perceive, remember, and think about. One of the most important objects for a primate is a face. Research on the macaque face patch system in recent years has given us a remarkable window into the detailed processes underlying object recognition. Here, we review the macaque face patch system, including its anatomical organization, coding principles, role in behavior, and interactions with other brain regions. We highlight not only how it constitutes an archetypal object recognition system, but also how it may provide a key to understanding mechanisms for higher cognitive function.

Introduction

The neural circuits underlying visual perception constitute one of the grand mysteries of neuroscience, spawning a huge number of discoveries concerning the development, structure, and function of the brain. More than a century ago, Sherrington marveled at the miracle of vision:

> The eye sends . . . into the cell-and-fibre forest of the brain throughout the waking day continual rhythmic streams of tiny, individually evanescent, electrical potentials . . . . A shower of little electrical leaks conjures up for me, when I look, the landscape; the castle on the height, or, when I look at him approaching, my friend's face (Sherrington, 1940).

How is this possible? We know that after entering the eye, visual information travels in the cortex through a sequence of retinotopic visual areas before entering a large brain region called inferotemporal (IT) cortex, the site of high-level object representation (Tanaka, 1996). Here, we review our current understanding of one particular system within IT cortex, the macaque face patch system, a network of regions dedicated to processing faces.

Why study the processing of faces? Wouldn't it be of greater interest to study "general" object recognition? Faces are indisputably a rather peculiar category of objects (see Box 1, "Are faces special?"). One thing that makes faces peculiar is that we are all experts at face recognition, able to distinguish two people based on the subtlest of differences in facial structure. And faces, unlike most objects around us, carry enormous social importance. For example, social psychologists have found that a split-second judgment of competence based solely on facial appearance could predict 69% of Senate races in 2004 (Todorov et al., 2005).

We focus our attention on the face patch system of the macaque monkey because we believe it constitutes the currently best understood system for high-level object representation. We will review its anatomical organization, coding principles, role in behavior, development, and interactions with other brain regions. In his famous book for young scientists, Sir Peter Medawar advised that a key to success in science is to find the "turtle's underbelly, the soft spot that makes a hard problem tractable" (Medawar, 2008). In this review, we argue that the macaque face patch system provides a useful model not only for understanding the neural mechanisms of high-level object representation, but also the myriad brain systems that operate on the output of these mechanisms including memory, thought, and action.

Face cell discovery

The discovery of face cells seems like one of those remarkable, unaccountable accidents of experimental science. How did it happen? As with most historical events, it seems that many forces contributed to this singular, serendipitous event. Charles Gross, as a young junior faculty member at MIT, was inspired by stories that he heard from his colleague Hans Lukas Teuber (Gross, 2006) about how temporal lobe lesions induced object agnosias including prosopagnosia. He was also stimulated by work on bug detectors in the frog’s retina and superior colliculus by his colleague Jerry Lettvin, as well as the famous work of Hubel and Wiesel across the river at Harvard Medical School on complex cells in area V1 and their speculation that even more complex cells would lie more anterior in the brain. He was also influenced by the Polish psychologist Konorski’s speculations on gnostic neurons (“grandmother cells” that encode highly specific concepts) (Konorski, 1967). And finally, there was a very relatable human element of seeking a new beginning after repeated failures. Gross recalls, “Discouraged by my inability to understand the frontal lobe, I decided it lay in an inaccessible limbo bearing little relationship to anatomy, physiology, and psychology…So I decided to turn my attention to the cortex on the inferior convexity of the temporal lobe: inferotemporal cortex.” (Gross, 2006) Here, we see the first glimmer of the turtle’s underbelly.

Gross and his colleagues reported the astonishing discovery of cells that responded selectively to various complex objects such as hands, faces, and trees, in both the upper and lower banks of the superior temporal sulcus (STS) (Gross et al., 1972; Bruce et al., 1981; Desimone et al., 1984). This finding was initially greeted with great skepticism by the neuroscience community. David Marr even commented, “Suppose, for example, that one actually found the apocryphal grandmother cell. Would that really tell us anything much at all?” (Marr, 1982). One major reason for the skepticism was that the cells seemed more or less randomly scattered across the temporal lobe, with some concentration in the STS (Baylis et al., 1985; Baylis et al., 1987), and it wasn’t clear how one might go about understanding them at a deeper level.

In 1997, Nancy Kanwisher and colleagues published a landmark paper reporting discovery of a face-selective area in the human brain using functional magnetic resonance imaging (fMRI) (Kanwisher et al., 1997) (this is the most-cited paper ever published in the *Journal of Neuroscience*). Several previous neuroimaging studies had strongly suggested the existence of such an area (Malach et al., 1995; Puce et al., 1995; Clark et al., 1996; Puce et al., 1996; Courtney et al., 1997; Haxby et al., 1999), including an early PET study by Sergent et al. from 1992(Sergent et al., 1992). The fMRI response of this area, which Kanwisher dubbed the "fusiform face area" (FFA), was about three times as strong to faces as to any other object. This area was located in the same place in every single subject. In addition to the FFA, several other face-selective regions were also found. The discovery of these face-selective regions in human ventral temporal cortex was remarkable because it suggested that face processing might be localized to specific chunks of cortex. If one could only record from these regions, perhaps one could understand how facial identity is represented. And this, in turn, could shed light on how other objects are represented, since the central computational challenge of face recognition, to recognize a face despite myriad "accidental" changes due to changes in lighting, viewing angle, partial occlusion, and so on, is the same as that of general object recognition (DiCarlo and Cox, 2007).

Various studies tried to deduce coding properties of the human FFA using fMRI. For example, in one approach, researchers used a method called fMRI adaptation to deduce that face-selective voxels are more invariant to changes in size and position than to changes in viewpoint and illumination (Sapountzis et al., 2010). *Ultimately, however, one cannot sort rice grains using a forklift*. The BOLD signal remains only an indirect and coarse measure of neural activity, and the response properties of single cells within face-selective regions remained unclear from human fMRI experiments.

Some studies even questioned the exclusivity of the FFA for face processing and suggested that it might actually be a region specialized for processing any category at which one is an expert (Gauthier et al., 1999). The key evidence in favor of the expertise hypothesis was that within the FFA, experts show higher activation to objects of their expertise compared to non-experts. In particular, it was found that bird experts show higher activation to birds, and greeble experts to greebles (a “greeble” is an artificial object category). But since both birds and greebles are somewhat face-like, this result could also indicate that the FFA codes objects fitting a coarse face template, rather than any object of expertise (with attentional modulation in experts). A subsequent study supported this viewpoint (de Beeck et al., 2006): expertise in discriminating three artificial categories (spikies, cubies, and smoothies) produced changes in distributed response patterns outside the FFA but not within the FFA, except in one subject who construed the smoothies as human-like (“women wearing hats”).

A path out of the confusion concerning what single cells in face-selective areas actually represent emerged with the discovery of face-selective regions in the macaque monkey using fMRI (Tsao et al., 2003a; Moeller et al., 2008) (see Box 2, “Face patches across species”). Comparing fMRI responses to images of faces vs. non-face objects revealed activation in several regions of IT cortex. This discovery made it possible to record electrophysiologically from single cells in fMRI-identified face patches and to systematically study the detailed selectivity of cells within these patches.

Early fMRI experiments found three face-selective regions (Tsao et al., 2003a), and later experiments using improved methods revealed approximately six face-selective regions (Moeller et al., 2008) in each hemisphere of macaque IT cortex (Fig. II-1): the posterior lateral (PL), middle lateral (ML), middle fundus (MF), anterior lateral (AL), anterior fundus (AF), and anterior medial (AM) patches. The finding of face-selective regions in the macaque with fMRI has been replicated by multiple groups (Pinsk et al., 2005; Pinsk et al., 2009; Issa and DiCarlo, 2012; Srihasam et al., 2012; McMahon et al., 2014; Afraz et al., 2015; McMahon et al., 2015; Aparicio et al., 2016; Arcaro and Livingstone, 2017).

Electrophysiological recordings targeted to face patches ML/MF revealed an astonishing phenomenon: almost all (97%) of the visually-responsive neurons in these two regions were strongly face selective (Fig. II-1c) (Tsao et al., 2006). Interestingly, in addition to a strong response to faces, many ML/MF cells showed a strong response to a clock and an apple; this gave a further hint that cells in these patches are coding faces and not arbitrary objects of expertise – even non-face objects effective in eliciting spikes were face-like. Subsequently, recordings in face patches PL (Issa and DiCarlo, 2012), AL (Freiwald and Tsao, 2010), AF(McMahon et al., 2014), and AM (Freiwald and Tsao, 2010) showed high concentrations of face cells as well (Fig. II-1c).

The discovery of such high concentrations of face cells was not at all expected based on previous electrophysiology in IT cortex (Baylis et al., 1985; Baylis et al., 1987) and suggested that the turtle's underbelly for object recognition might be at hand. A large number of detailed questions about the neural mechanism for object recognition suddenly appeared tractable. What are cells in each patch coding? What are the functional differences between patches? What is the connectivity of these specialized patches to each other and the rest of the brain? How does activity in each patch contribute to perception and behavior? Is the invariance problem for faces solved by this system of patches and if so, how? Is this system of patches unique to faces or does it generalize across all of IT cortex?

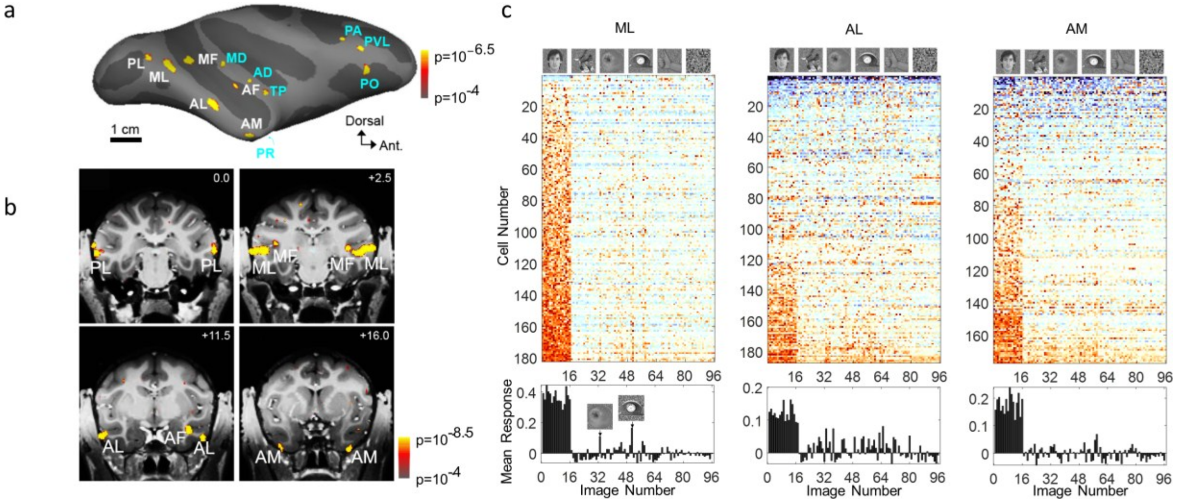


Figure II-1: Face-selective patches in macaque cortex. *(**a**) Schematic of patches shown on inflated right hemisphere of a macaque. The canonical six in IT cortex are*

*labeled in white, while additional patches are labeled in cyan. (**b**) Patches shown on coronal slices, mm relative to inter-aural canal indicated in upper right corner. (**c**) Top: Response profiles of all visually responsive cells recorded from ML, AL, and AM, to 96 pictures, consisting of faces, human headless bodies, fruits, gadgets, hands, and scrambled patterns (left to right, 16 images/category). Stimuli were presented for 200 ms and separated by 200 ms inter-stimulus intervals. The two non-face stimuli with the highest response in ML are shown (clock and apple). Bottom: Normalized population response to each image (averaged across all units). For each cell, responses were summed from 100 to 300 ms, averaged over all presentations, baseline subtracted, and normalized to the maximum average response. Panels (a) & (b) adapted from (Moeller et al., 2008) panel (c) (left) from (Tsao et al., 2006), and panel (c) (middle, right) from (Freiwald and Tsao, 2010).*

Anatomical organization

Spatial organization of face patches*.* The face patches are located within cytoarchitectonic areas TEO (PL) and TE (ML, MF, AL, AF, AM), the posterior and anterior part of IT cortex, respectively (Fig. II-1a, II-1b). The six patches are located symmetrically in the two hemispheres and in similar locations across different monkeys (Tsao et al., 2003a; Tsao et al., 2008a). This provided one of the first hints that the patches might constitute a system for face processing, rather than being simply random islands of face-selective cortex. There is individual variability both in the number and position of patches. Some animals show only a subset of the typical pattern of six bilateral patches, while some show even more than the canonical six (Tsao et al., 2008a). In many animals, patches PL and ML are confluent, making it hard to distinguish the two. In some animals, AL and AM are also confluent.

In addition to this canonical pattern of six patches, several additional patches activated by the contrast faces versus objects can be found in the temporal lobe, in the upper bank of the STS (middle dorsal patch MD and anterior dorsal patch

AD) (Fisher and Freiwald, 2015), perirhinal cortex (PR), and the temporal pole (TP) (Landi and Freiwald, 2017) (Fig. II-1a). These patches are more variable across animals. Ku et al. reported further additional face patches in entorhinal cortex, parahippocampal cortex, amygdala, ventral TE, and hippocampus, using a high-field spin-echo sequence (Ku et al., 2011).

Face patches reside not only in the temporal lobe. Within prefrontal cortex, there are three regions of face-selective cortex in ventral prefrontal cortex: PO (prefrontal orbital), PA (prefrontal pre-arcuate), and PVL (prefrontal ventrolateral) (Tsao et al., 2008b) (Fig. II-1a).

Anatomical connectivity of face patches*.* A large number of anatomical studies have probed the connectivity of IT cortex. These studies have generally found patchy connectivity within IT, hinting at the presence of modular networks (Saleem et al., 1993; Borra et al., 2010). However, most of these studies have been conducted independently of functional testing in the same animals. *It is hard to learn about the life of a snake by observing its long abandoned coil.* The study of face patches has introduced a new paradigm for investigating how functional properties and anatomical connectivity relate, leveraging the ability to target tracers to anatomical regions defined by functional selectivity.

To map the connectivity of macaque face patches, Moeller et al. electrically microstimulated specific face patches while performing simultaneous fMRI (Moeller et al., 2008). Stimulation of each of five targeted face patches produced strong activation in other face patches. For example, stimulation of ML produced activation in PL, ML, AL, AF, and AM (Fig. II-2a). Activation was largely ipsilateral. Within IT cortex, activation elicited by face patch stimulation was almost entirely confined within other face patches. Another study that directly compared face patch stimulation and body patch stimulation found that there was very little overlap between functional connectivity patterns of the two, suggesting that the face patch network and body patch network are largely parallel systems (Premereur et al., 2016).

Stimulation of IT face patches also activated several subcortical regions, including the amygdala, claustrum, and pulvinar (Moeller et al., 2008). In particular, stimulation of AM elicited activation in the amygdala and claustrum, while stimulation of ML and AL both activated the inferior pulvinar (Moeller et al., 2008) (Fig. II-2b). Because the pulvinar, claustrum, and amygdala were the only brain structures consistently activated by stimulation of the face patches outside the face patch network, these structures may constitute three hubs by which face patches communicate with other regions of the brain.

More recently, direct injections of retrograde tracers have been made into four face patches: PL, ML, AL, and AM (Grimaldi et al., 2016). Grimaldi et al. found strong inputs to the most posterior face patch PL from specific regions in area V4 and TEO, and V2 and V3 more weakly (Grimaldi et al., 2016). Within IT cortex, anatomical tracing experiments largely confirmed electrical microstimulation experiments: the IT inputs to a particular face patch largely came from other face patches (Fig. II-2c).

It is generally thought that processing increases in complexity from posterior to anterior in the temporal lobe (Tanaka, 1996), raising the question whether there is a hierarchy between face patches. Grimaldi et al. did not find anatomical evidence for a strict hierarchy (despite functional evidence for this, see "Coding principles" below). For example, a retrograde tracer injection into AM, the most anterior face patch, revealed inputs from both AL and ML (Fig. II-2c, right). And in the same animal, when ML was injected, inputs were found from both AL and AM (Fig. II-2c, left). Thus there is extensive feedback within the system, and each patch talks to multiple other patches (Fig. II-2d). Furthermore, the neurons projecting from one patch to another were located in both supra and infragranular layers, consistent with a bidirectional hierarchical relationship (Felleman and Van Essen, 1991).

The major inputs from higher order areas to face patches as revealed by tracers came from the medial temporal lobe as well as three subcortical regions, the claustrum, pulvinar, and amygdala, largely confirming microstimulation results. Claustrum cells were strongly clustered and segregated by their target face patch. Projections from prefrontal cortex were extremely weak; this is very surprising in light of previous retrograde injections in IT showing strong prefrontal input (Borra et al., 2010; Saleem et al., 2014), as well as results of electrical microstimulation showing that stimulation of AM activates prefrontal face patches PO, PA, and PVL (Moeller et al., 2008). This discrepancy between results of electrical stimulation and anatomical tracing may indicate that the connection from IT to prefrontal face patches is not direct.

Interestingly, one animal in this tracer study showed an extra patch at an unusual location, and this patch also showed highly specific anatomical connections with other face patches in IT cortex (Fig. II-2e), suggesting it was a full-fledged node of this animal’s face processing system (Grimaldi et al., 2016). This underscores that individual variability between animals is real and not due to noise, and could explain findings of “extra” clusters of face-selective cortex outside the canonical six in electrophysiological and imaging experiments.

Overall, the specificity of connections between face patches suggests that they constitute a specialized system for face processing, and the small number of connected nodes outside the face patch system suggests that the output of the system is read out by a small number of downstream nodes.

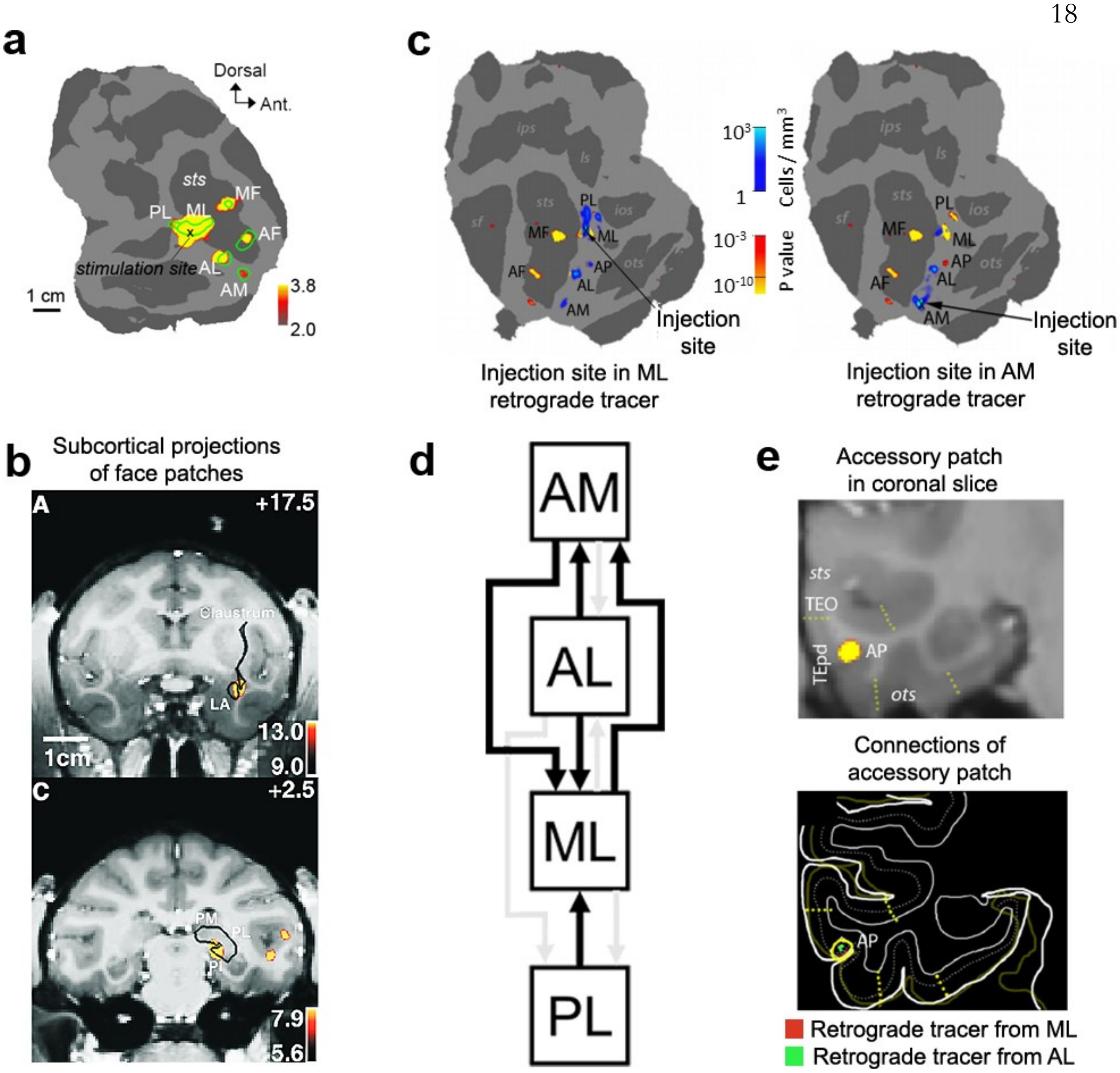

a
Dorsal
Ant.
sts
MF
PL
ML
AF
stimulation site
AL
AM
1 cm
3.8
2.0
c
ips
ls
sts
sf
ios
ots
PL
MF
ML
AP
AF
AL
AM
Injection site
10^3
1
Cells / mm^3
10^-3
10^-10
P value
Injection site in ML retrograde tracer
Injection site in AM retrograde tracer
b
Subcortical projections of face patches
A
+17.5
Claustrum
LA
1cm
13.0
9.0
C
+2.5
PM
PL
PI
7.9
5.6
d
AM
AL
ML
PL
e
Accessory patch in coronal slice
sts
TEO
TEpd
AP
ots
Connections of accessory patch
AP
🟥 Retrograde tracer from ML
🟩 Retrograde tracer from AL

Figure* II-2*: Anatomical connectivity of face patches. *(**a**) Connectivity of IT face patches revealed by microstimulation combined with fMRI. The electrode targeted ML in the right hemisphere. Areas significantly activated by microstimulation of ML are shown overlaid on a flatmap. (**b**) Microstimulation of face patch AM (top) and ML (bottom) activates the amygdala (LA: lateral amygdala), claustrum, and pulvinar (PI: inferior pulvinar; PM: medial pulvinar; PL: lateral pulvinar). (**c**) Left: Retrograde injections into ML revealed labeled cells in PL, AL, and AM; right: retrograde injections into AM revealed labeled cells in ML and AL. Blue indicates labeled cells, yellow indicates fMRI-identified face patches. (**d**) Wiring diagram schematic of face* *patches. Black arrows indicate results from the tracer injections into ML and AM in (c). Grey arrows indicate further results from tracer injections into PL and AL not shown in (c). (**e**) Top: Face patch AP in coronal slice. Bottom: Tracer injections in ML (red) and AL (green) both revealed labeled cell bodies in AP. Panels (a) and (b) adapted from (Moeller et al., 2008), panels (c), (d) and (e) adapted from (Moeller et al., 2008; Grimaldi et al., 2016).*

Sub-compartmentalization within patches. Do face patches constitute atoms of IT cortex, or can they be further subdivided into finer modules? The size of face patches is only roughly estimated by fMRI experiments and depends on the threshold. We mapped the extent of face patch ML using both fMRI and electrophysiology, and found that the size of fMRI-defined ML (at threshold of $p<10^{-10}$) was ~4 mm in diameter, while the electrophysiology-defined patch was ~2 mm in diameter, with a sharp border going from all to no face-selective cells (Grimaldi and Tsao, unpublished). In contrast, Aparicio et al. reported that the maximum diameter of face patch ML as mapped with electrophysiological recordings was 6 mm, i.e., similar to its fMRI activation, with the fraction of face-selective cells decreasing monotonically from its center (Aparicio et al., 2016).

To determine whether there is finer structure within a face patch, Sato et al. performed optical imaging to identify face-selective regions in the vicinity of AL and then performed single-unit recordings in multiple sites within identified face-selective regions (Sato et al., 2013). They found heterogeneity in selectivity for human versus monkey faces and scrambled versus intact faces across different sites, but similar selectivity within each site, suggesting the existence of finer feature columns (~0.5 mm in diameter) within face patch AL.

Rajimehr et al. used retinotopic face stimuli to map face patches in an fMRI experiment and found heterogeneous retinotopic organization within face patches (Rajimehr et al., 2014). For example, ML/MF was subdivided into a posterior-ventral subregion that showed variations in eccentricity and polar angle selectivity and an anterior-dorsal subregion with no retinotopic selectivity. Using high-resolution fMRI, Janssens et al. mapped retinotopy across IT and found that ML showed a foveal bias, while more anterior patches showed no retinotopic organization (Janssens et al., 2014).

Further insight into clustering of feature selectivity within face patches comes from a multi-voxel pattern analysis (MVPA) study that used activity from fMRI voxels to decode specific features of faces that monkeys viewed (Dubois et al., 2015). The authors found that within single face patches, different voxels were tuned to different viewpoints of a face, and the viewpoint of a face could be decoded from the fMRI response pattern in face patches ML/MF. On the other hand, face identity could not be decoded from the voxel response pattern. This suggests that viewpoint, but not identity, shows clustering of tuning on the coarse scale of fMRI voxels (~1 mm). Ultimately, higher resolution techniques for functional mapping, e.g., two-photon calcium imaging (Tang et al., 2018b), will be needed to more clearly understand sub-compartmentalization in face patches and to clarify whether the border between face patches and neighboring IT regions is gradual or discrete.

Coding principles

What are cells in face patches coding? In general terms, cells could be detecting faces, identifying faces, or recognizing other aspects of faces such as expression, dominance, or gender (Bruce and Young, 1986). The simple fact that almost all cells within a face patch are face-selective suggests that face detection is one major computational goal. But even a cursory glance at responses across a population of ML/MF cells to 16 different faces shows that individual cells respond differently to different faces (Tsao et al., 2006) (Fig. II-1c). An early study reported that face-selective cells in IT encode the physical structure of faces while cells in the anterior superior temporal polysensory area encode familiarity of faces (Young and Yamane, 1992). However, lack of access to face patches precluded a detailed understanding.

In this section, we summarize the different aspects of faces that are coded by cells in face patches.

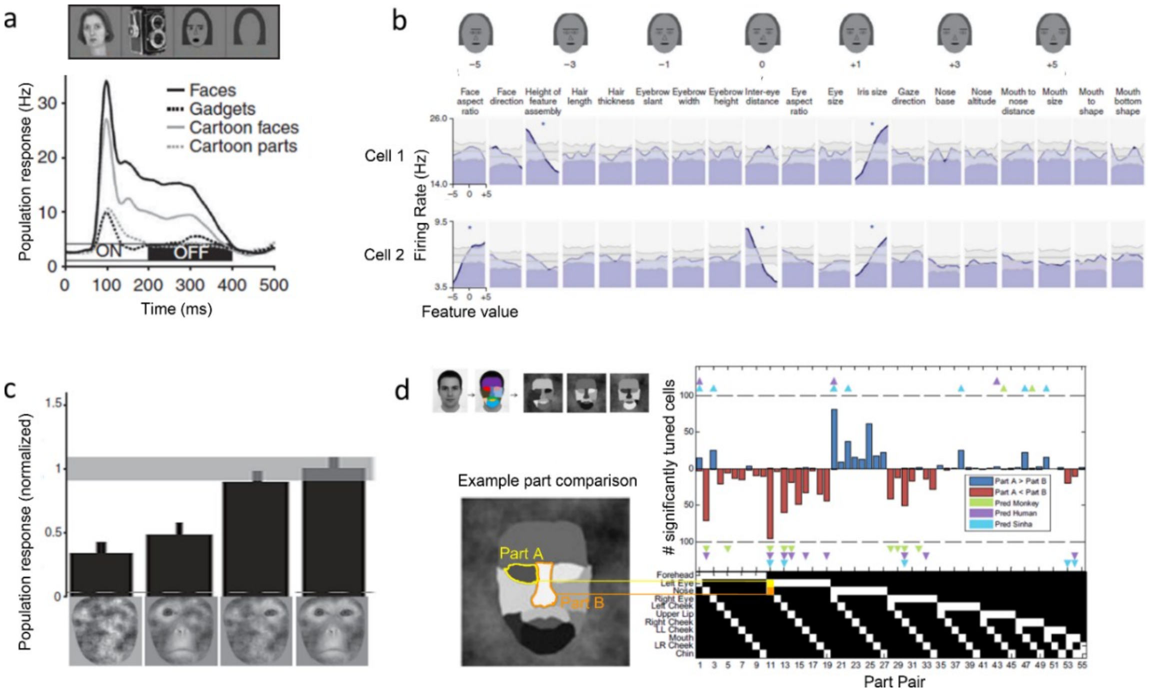


Figure* II-3*: Probing feature selectivity of face cells. *(**a**) Mean response time course from 66 ML/MF cells to four stimulus categories: faces, gadgets, cartoons, and cartoon parts. (**b**) Top: depiction of cartoon stimuli varied across inter-eye distance. Bottom: Two example cells show ramp-shaped tuning. Asterisks mark significant modulation (P < 0.001). (**c**) Population response (110 cells) to images containing some or all of the face parts in the context of the face outline. (**d**) Significant contrast feature histogram showing consistency of selectivity for contrast relationships across the population. Blue (red) bars indicate the number of cells significantly preferring the intensity of part A to be greater (less) than the intensity of part B. Triangles indicate feature pairs predicted to elicit significant contrast polarity tuning across the population by three different models, together with the predicted direction of tuning. Pred monkey: A monkey face was illuminated from different directions and the contrast relationships that were robust to changes in illumination were identified. Pred human: the same was done for a human face. Pred Sinha: similar to Pred human, reported in a different study(Sinha, 2002). Panels (a) and (b)* *adapted from (Freiwald et al., 2009), panel (c) from (Issa and DiCarlo, 2012), panel (d) from (Ohayon et al., 2012).*

Probing face coding in ML/MF using cartoon faces*.* Early experiments tried to probe mechanisms for face detection, face identification, and holistic processing using cartoon faces (Freiwald et al., 2009). The motivation for using cartoons was twofold. First, cartoons are able to convey rich information about not only the presence of a face, but also individual identity and expression. Second, a much smaller set of parameters is required to specify a cartoon face compared to a real face.

To address the mechanism for face detection, cartoon faces were constructed using composites made of 7 cartoon parts (face outline, irises, mouth, nose, hair, eye, eyebrow), resulting in 2^7 possible combinations of parts being either present or absent. Cells in ML/MF responded strongly to the cartoon faces, with the mean response to cartoon faces being ~80% that to photographs of real faces (Fig. II-3a). Single cells were selective for the presence of specific subsets of parts. Half of the response variance (52%) could be explained by linear regression on presence or absence of the 7 face features. An additional 18% of the variance could be explained by second order effects. This dependence of responses on multiple parts and part interactions shows that middle face patch neurons are not just detecting isolated features. However, because 70% of the response variance could be explained by first and second order effects alone, middle face patch cells are not highly nonlinear holistic cells either.

To address the mechanism for face identification, 19 features defining a cartoon face (including inter-eye distance, face aspect ratio, iris size, etc.) were parametrically varied. Single cells showed ramp-shaped tuning to subsets of face features, i.e., a monotonically increasing response from one feature extreme to the opposite extreme (Fig. II-3b) (Leopold et al., 2006).

To address the mechanism for holistic processing (see Box 1, “Are faces special?”), the effect of context on face feature tuning was examined (Freiwald et al., 2009). When tuning to a particular feature (e.g., face aspect ratio) was measured while other features were present, the gain of tuning was on average twice that when the same feature was shown in isolation. This provides a mechanistic explanation for the whole-part effect (Tanaka and Farah, 1993). Is the gain increase due to simple proximity of other facial features, or does their spatial arrangement also matter? Measurement of tuning to features in both upright and inverted faces revealed that cells were tuned to 25% fewer features when faces were inverted. Furthermore, substantial tuning to mouths appeared. This result suggests that mere proximity of other features alone is insufficient: face cells match the incoming feature constellation to an upright face template, and interpret features in the context of this template (thus, mouths may become interpreted as eyes in some cells).

Issa and DiCarlo selectively unveiled different parts of a natural face to ask which parts triggers cells to fire (Issa and DiCarlo, 2012). They targeted their experiments to the most posterior face patch PL, reasoning that this should reveal the earliest template used by the brain for face detection. They found that the presence of a single eye surrounded by a face outline could drive a significant response in nearly all PL cells. Indeed, images containing only an eye evoked almost the same response as full-face images (Fig. II-3c).

In natural faces, unlike cartoon faces, diagnostic shape elements (eyes, nose, etc.) occur with stereotyped contrast relationships (e.g., the eyes are usually darker than the nose). Pawan Sinha has suggested that the brain could detect faces by pooling contrast features, with the relevant contrast features being ones that are invariant to changes in lighting (Sinha, 2002). A similar scheme was used by one of the most powerful early algorithms for face detection, the Viola-Jones algorithm (Viola and Jones, 2001).

To investigate whether face cells exploit contrast features as well, Ohayon et al. presented artificial face stimuli consisting of 11 different regions varying in brightness (Ohayon et al., 2012). In half the cells, activity could be driven from no response to a response greater than that to a real face by changing contrast combinations. Cells were highly consistent in their contrast polarity preferences (Fig. II-3d). Moreover, these preferences matched features used by a computer vision algorithm for face detection. Finally, most cells showed tuning to both contrast polarity and geometry of facial features, suggesting single cells contribute to both detection and recognition.

Overall, experiments with simplified face stimuli reveal the exquisite precision with which face cells measure physical properties of faces, including which features are present, their geometry and contrast, and surrounding context.

Cells show increasing view invariance in more anterior patches. One of the central computational challenges of object recognition is to recognize an object's identity despite view, partial occlusion, lighting, and other "accidental" changes. Freiwald and Tsao presented 25 different identities each at 8 different head orientations while recording face cells from multiple patches, and discovered a striking functional distinction between patches in terms of their view tuning (Freiwald and Tsao, 2010). In ML and MF, neurons respond selectively to specific views. In AL, there are two populations of neurons that respond similarly to left and right profile views, and to up/down/frontal views, respectively, thereby achieving partial view invariance. In AM, the most anterior patch, neurons respond invariantly across all views (Fig. II-4a). Some cells in AM responded extremely sparsely to only a small subset of face identities, with invariance across changes in view (Fig. II-4b). A major purpose of the face patches may thus be to construct a representation of individual identity invariant to view direction. Furthermore, cells not only increase in view invariance from ML to AM, but also in size and position invariance (Fig. II-4c).

The existence of cells showing mirror symmetric identity selectivity in AL is puzzling. It makes sense that there should exist a patch where cells have a view-specific representation, since the actual visual input is view specific. And it makes sense that there should exist a patch where cells have a view-invariant representation, since this is the end goal; however, a representation where the responses to left and right views are the same is surprising, and various attempts have been made to explain this computationally. In one study, Leibo et al. showed that a feedforward network trained on sequences of images of faces rotating from left to right using Hebbian learning develops a mirror-symmetric representation of faces in an intermediate layer and a view-invariant representation in a later layer (Leibo et al., 2017), just as observed in the face patch system. In another study, using a model that inverts a generative 3D face graphics engine, Yildirim et al. were also able to replicate the three stages of increasing view invariance observed in the face patch system (Yildirim et al., 2018). A definitive explanation for why Nature has broken up the problem of invariant face representation into this particular set of steps (ML/MF -> AL -> AM) remains elusive. Furthermore, the actual sequence of steps leading to a view-invariant representation in AM may be more complex, given that the anatomical connectivity is inconsistent with a strict hierarchy (e.g., there is a direct projection from the middle face patch ML to the most anterior face patch AM, Fig. II-2c).

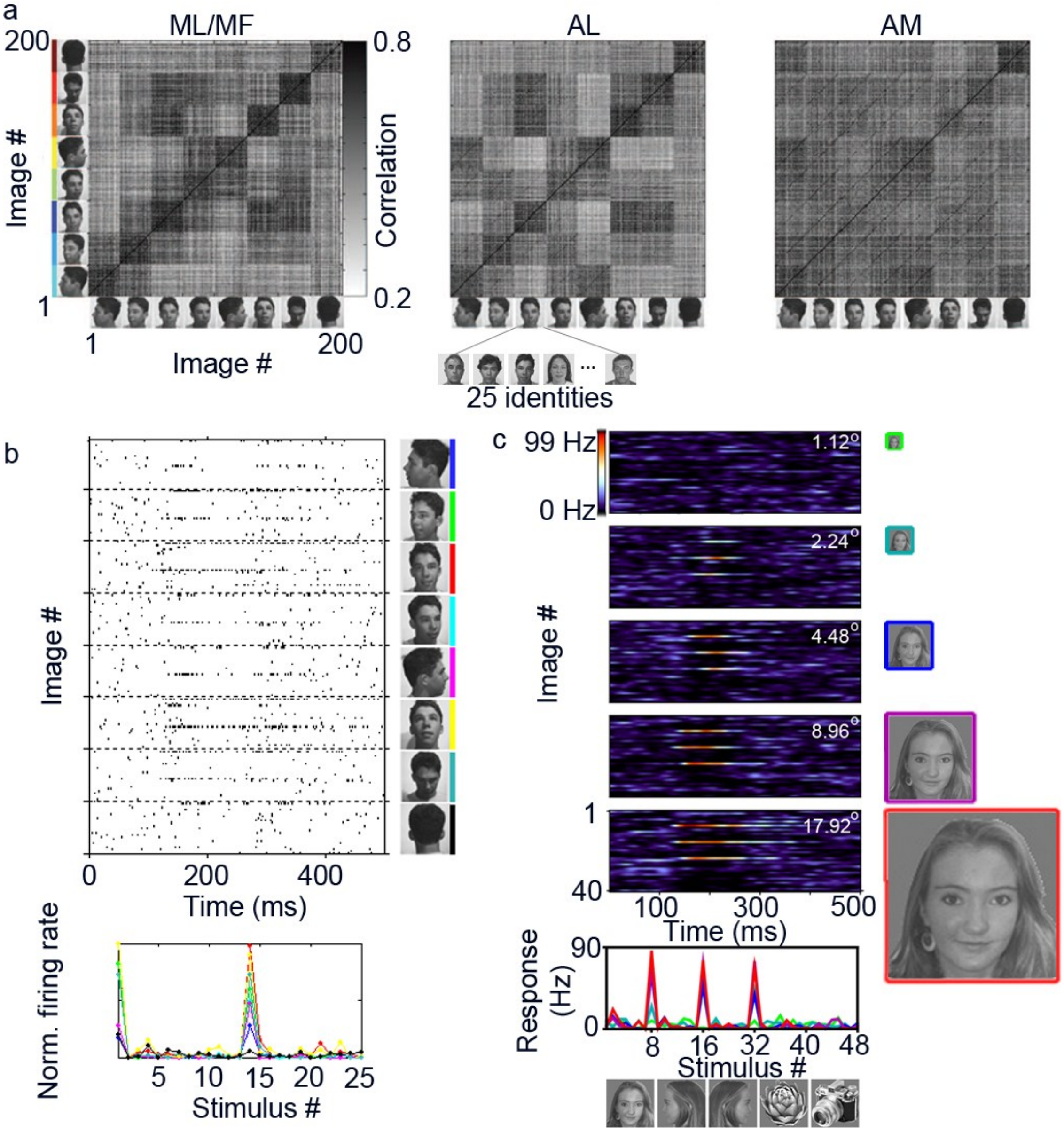


Figure* II*-4: Increasing view invariance across the face patch hierarchy. *(**a**) Representation similarity matrices computed from responses to the 200 face view stimuli for ML/MF, AL, and AM. (**b**) Top: Mean response time course of an example AM cell that responded very sparsely to the 200 face view stimuli. Bottom: Mean response to 25 individuals at 8 views. (**c**) Mean response time courses of an example AM cell to a set of faces and objects, each at five different sizes (stimulus size indicated in upper right). (Middle) Mean responses to the 40 images, with each color indicating one size. (Bottom) Examples images from the 40 image set. Only 3* *of the 40 images elicited responses from this cell, corresponding to the face of one individual regardless of view and size. Adapted from (Freiwald and Tsao, 2010).*

Cracking the code for facial identity*.* Finding robust tuning of neurons along some stimulus dimensions (e.g., identity, viewpoint) falls short of fully understanding the code of the neurons. If one could reconstruct an arbitrary real face seen by a monkey using responses of face cells and predict a neuron's response to an arbitrary real face, then one could claim to truly understand the code for facial identity.

To discover the brain's code for realistic facial identity, Chang and Tsao parameterized a large database of faces using 50 features per face (Chang and Tsao, 2017), adopting an approach to generating faces originally developed in computer vision called the "active appearance model" (Cootes et al., 1998; Chang and Tsao, 2017). Specifically, for each face in the database, they placed a set of landmarks on key features by hand. The (x, y) coordinates of these landmarks describe the "shape" of the face. They then morphed these landmarks to match those in a standard template. The resulting image describes the "shape-free appearance" of the face (Fig. II-5a). Finally, they performed principal components analysis independently on the shape vectors and appearance vectors across the entire set of faces. By taking the top 25 shape principal components and the top 25 appearance principal components, they created a 50 dimensional face space (Fig. II-5b).

They randomly drew 2000 faces from the 50-d face space and presented them to the monkey while recording cells from two face patches, ML/MF and AM. Single cells showed ramp-shaped tuning to a subset of the 50 face parameters (Figure II-5c). ML/MF cells were preferentially tuned to shape parameters while AM cells were preferentially tuned to appearance parameters, consistent with the increased invariance to view of AM cells found earlier (Freiwald and Tsao, 2010).

The finding of ramp-shaped tuning to subsets of face features suggests that the response of each cell can be roughly approximated by a linear combination of facial features. Conversely, the 50 face features should be a linear transform of the population response vector. Importantly, because the experimenters presented the same set of 2000 faces to each cell, they could directly test this. Reconstructions of faces using linear regression on activity of just 205 face cells from ML/MF and AM were strikingly accurate (Fig. II-5d).

If a neuron's response can be completely explained by linear regression, this implies a very simple geometric picture of the underlying computation: the neuron is projecting incoming faces onto a specific preferred axis and measuring the value along that axis. This has the surprising consequence that the neuron should respond exactly the same to all faces in the hyperplane orthogonal to the preferred axis (Fig. II-5e). Note that this is not a logical necessity of ramp-shaped tuning (Fig. II-5f). To test this prediction, Chang and Tsao generated a series of faces with very different identity whose variation was orthogonal to the preferred axis of the cell. All of these faces turned out to elicit exactly the same response (Fig. II-5g).

In sum, for faces, the brain's code fortuitously turns out to coincide with a classic approach to parametrizing faces from computer vision: shape-appearance coordinates.

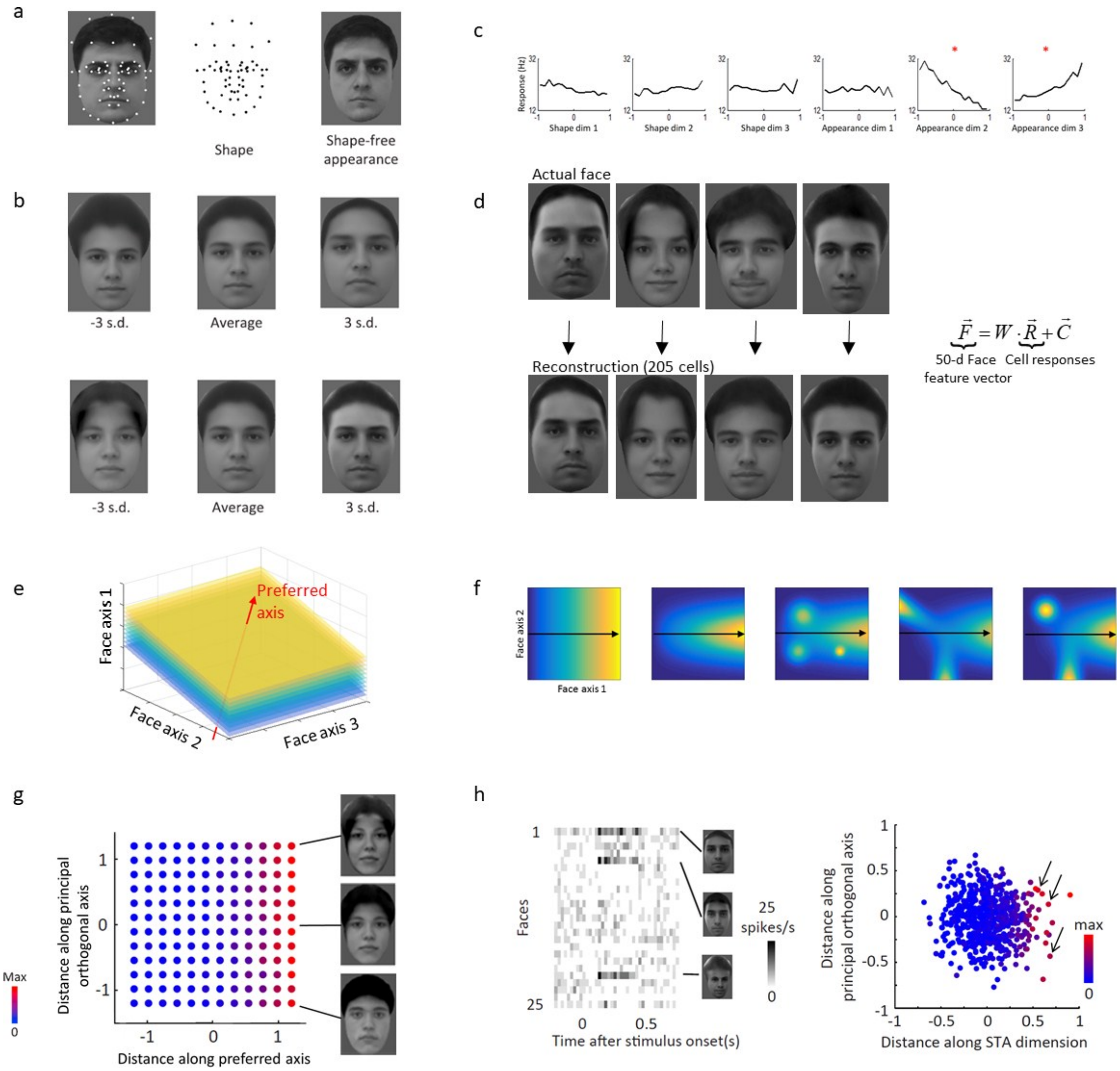


Figure* II*-5: The neural code for facial identity. *(**a**) Parameterizing faces: for 200 face images, 58 landmark points were annotated, as in the example shown (left). The positions of these landmarks describe the shape of the face (middle). The landmarks were morphed to an average template, resulting in an image describing the shape-free appearance of the face (right). (**b**) The first principal component for shape (top) and appearance (bottom). (**c**) Tuning to first three shape and first three appearance dimensions for an example neuron. Asterisks indicate neurons showing significant tuning. (**d**) Facial images reconstructed using facial features decoded by linear regression from responses of 205 cells in ML/MF and AM. Predicted faces* *and the corresponding actual stimuli presented in the experiment are shown. (**e**) Geometric picture of essential computation performed by a face cell: projection of incoming faces, formatted as shape appearance coordinates, onto a specific preferred axis. (**f**) Ramp-shaped tuning along a preferred axis does not imply flat tuning along orthogonal axes. Different examples of 2D tuning functions are shown. While all exhibit ramp-shaped tuning along face axis 1, only the leftmost example shows flat tuning along the orthogonal axis. (**g**) The responses of an AM cell to 144 faces evenly sampled from the 2D space spanned by the preferred axis and principal orthogonal axis, synthesized specifically for this cell, are color coded and plotted. (**h**) Left: responses of a sparse AM cells to 25 identities reveals strong responses to three disparate identities. Right: The response of this cell to 2000 faces reveals axis tuning. The three effective identities from the left plot are indicated by arrows: their difference lies on the null space of this cell. Adapted from (Chang and Tsao, 2017).*

From faces to objects*.* Is the mechanism used by the brain to encode facial identity, through a set of patches with increasing view invariance that collectively represent faces by projection onto a set of axes spanning face space, unique to faces? Bao et al. (Bao et al., 2019) recently explored the overall large-scale organization of IT cortex using very similar methods to those used to decipher the face patch system: fMRI, fMRI-guided electrophysiology, electrical microstimulation, and mapping of response selectivity within a computationally-generated object space framework. Using a deep network, they built a parametric object space by computing the first 50 principal components of responses of units in a deep neural network (AlexNet) to a large set of objects (Fig. II-6a). They then measured responses of monkey IT cells to dimensions of this space and found that cells are clustered to form a coarse topographic map of the first two dimensions of this space (Fig. II-6b). Furthermore, this map is replicated three times, with increasing view invariance (Fig. II-6c). In particular, the map consists of four quadrants encompassing face patches, body patches, as well as two newly discovered networks, one representing “spiky” objects and another representing “stubby” objects. Electrical microstimulation combined with simultaneous fMRI revealed that patches within the two newly discovered networks are anatomically connected, just like the face (Moeller et al., 2008) and body (Premereur et al., 2016) networks. Cells in each network approximately project incoming objects, formatted as vectors in object space, onto specific preferred axes, with the sign of the first two components of this axis dictated by the cell's anatomical position in IT cortex.

This work shows that IT cortex contains multiple siblings of the face patch network, which each share its hallmark properties of modular connectivity, functional hierarchy, and axis coding. Furthermore, the sibling networks are arranged in a way that is predicted by the topography of object space (Fig. II-6c). Thus a new general principle of IT cortex organization, "mapping object space," can explain both category and non-category selective regions. Furthermore, arbitrary objects could be reconstructed using responses of cells from this IT object-topic map (Fig. II-6d), suggesting that cells in this map provide a close-to-complete basis for general object space. Thus within IT cortex, the original promise of the face patches, to provide a turtle's underbelly for understanding the brain more generally, has been powerfully delivered.

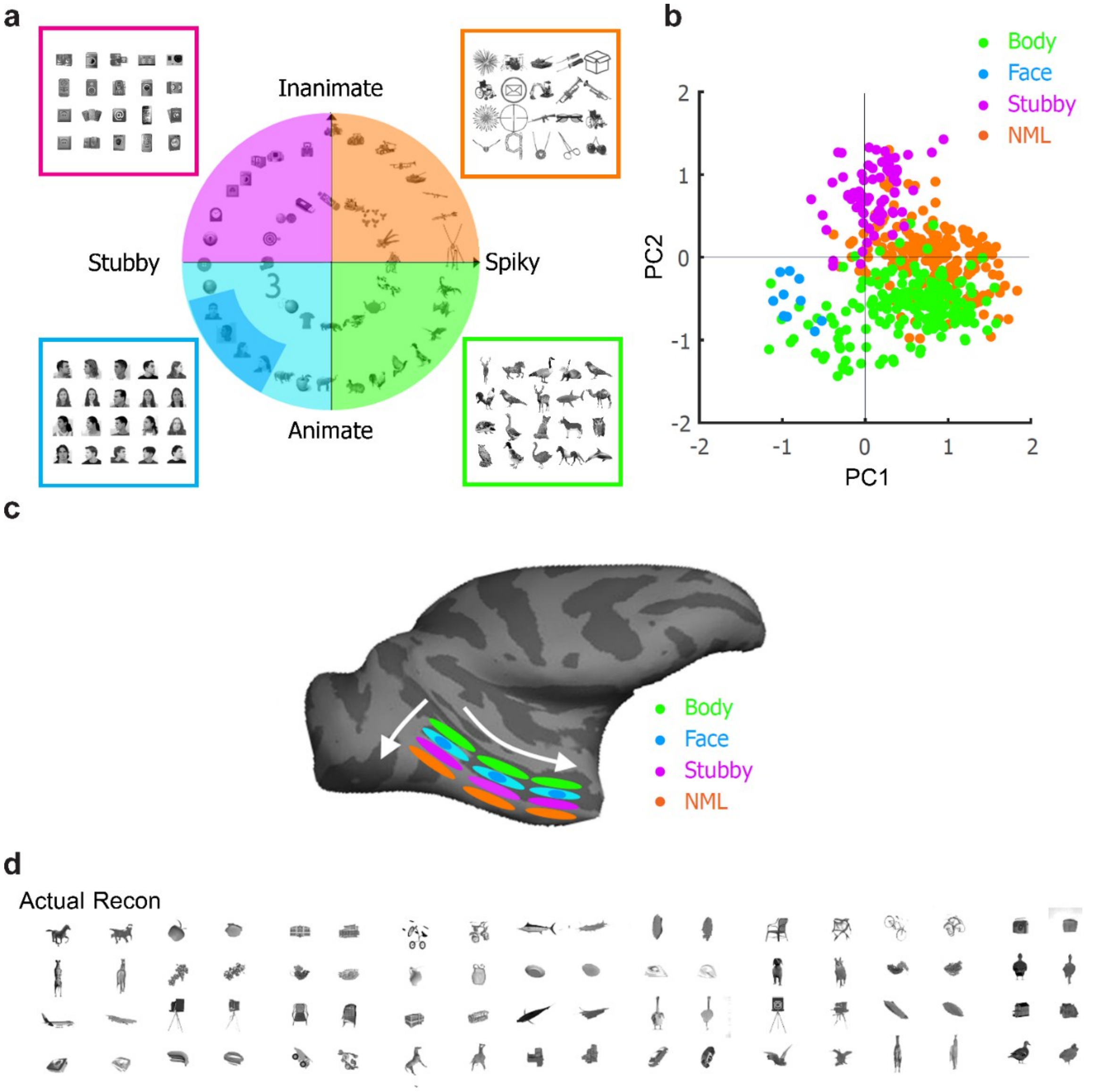


Figure* II-6*: The neural code for object identity in general. *(**a**) Schematic showing the four quadrants spanned by the first two PCs of object space. The stimuli in each of the quadrants were used for mapping four networks (face patches, body patches, NML patches, and Stubby patches) using fMRI. (**b**) Projection of preferred axis of each cell (N = 482) onto PC1 versus PC2 for all neurons recorded across four networks (NML network: yellow, body network: green, face network: blue, stubby network: magenta). (**c**) Schematic showing the threefold-repeated topographic map in IT that is organized according to the four quadrants of object space. (**d**) Example reconstructed images. Random objects were reconstructed using activity of 482* *cells sampled from the four networks and a generative adversarial network. Adapted from (Bao et al., 2019).*

Selectivity for dimensions beyond facial identity*.* Besides shape and appearance dimensions, other face-related properties such as expression, familiarity, and gaze have been found to be encoded inside and outside the canonical face patch system.

Facial expression is an extremely important feature for social interactions, and appears to be explicitly extracted by regions outside the canonical IT face patches, including the amygdala, orbitofrontal cortex, and motion-selective areas in the STS. FMRI experiments comparing expressive vs. neutral faces show only weak modulation in canonical IT face patches, but strong modulation in prefrontal face patch PO in the lateral orbital sulcus (Tsao et al., 2008b). Consistent with this, single-unit study targeting face cells in the lateral orbital sulcus, where PO is typically located, reveal many cells modulated by facial expression as well as other social dimensions such as gender and age (Barat et al., 2018). An fMRI study in monkeys (Hadj-Bouziane et al., 2008) reported that both IT and amygdala have maps of valence of facial expressions, and these valence maps overlap but don’t coincide with the canonical face patches. Further supporting an anatomical dissociation between processing of facial expression vs. identity, multivariate pattern analysis applied to monkey fMRI data revealed that both dynamic and static facial expression can be read out better from motion-selective areas in the STS than from face-selective areas (Furl et al., 2012).

Gaze is another important feature for inferring intent of other social agents. Using fMRI, Marciniak et al. found a patch in TEO that is activated when monkeys follow the gaze direction of a presented face (Marciniak et al., 2014). This patch is not activated when monkeys are trained to saccade based on identity rather than gaze direction. In face patch AD located in the upper bank of the STS, She and Tsao found cells that respond to facial movements, in particular change of gaze (She and Tsao, 2017). Behaviorally, Roy et al. found that inactivation of posterior STS causes an impairment of gaze following in monkeys (Roy et al., 2012). These results suggest the presence of modules separate from the canonical face patches that infer the gaze target of other agents.

One of the most important functions of face recognition is to identify familiar individuals. Landi and Freiwald reported that face patches PR and TP in the medial temporal lobe show more fMRI activation to faces that a monkey is personally familiar with than unfamiliar faces (Landi and Freiwald, 2017). She and Tsao recorded from PR and AM and found that in both regions, cells use axis codes for representing identity which are modulated by familiarity (She, 2018).

To study face patches in a more naturalistic setting, researchers have probed responses of face patches to natural movies. Park et al. showed natural movies to monkeys during both single-unit recordings and while performing fMRI. They found that even within a small volume of a few hundred microns in face patch AF, there is a large diversity in activity profiles of individual cells, with only a small subpopulation (~16%) of cells showing a time course resembling their local voxel or LFP signal (Park et al., 2017). In particular, McMahon et al. found that neural populations in AF are tuned to the size of a face in a natural movie, which correlates with the social variable of how close it is to the observer (McMahon et al., 2015). Fisher and Freiwald reported that patch MD in the upper bank of the STS is activated by natural movies of moving faces but responds much less to static faces or incoherent movies in which frames are presented in random order (Fisher and Freiwald, 2015).

At the other end of the spectrum, selectivity of face cells for low-level feature statistics has also been probed. Face cells in IT are tuned to spatial frequency invariant to distance of the face, unlike face cells in the amygdala, which are tuned to spatial frequency in retinal coordinates (Inagaki and Fujita, 2011). A study probing selectivity for the orientation content of faces found that ML cells are preferentially tuned to horizontally-filtered faces while AL cells are preferentially tuned to vertically-filtered faces (Taubert et al., 2016).

Coding of multiple objects. In the real world, primates need to recognize objects in the presence of other objects, and this presents a major coding challenge for IT, since IT cells have large receptive fields that typically encompass multiple objects. Bao and Tsao studied how cells in two category-selective regions of the macaque, one of them a face patch, integrate multiple objects. They found that cells in face and body patches consistently perform winner-take-all under certain conditions (Bao and Tsao, 2018). For example, when a face is presented above a non-face object, face cells respond as if only the face is present. They showed that this winner-take-all behavior could be parsimoniously explained by normalization within a pool of clustered face cells. Indeed, accomplishing winner-take-all, which can be construed as a form of hard-wired attention, could be one important evolutionary driving force for modularity in IT.

Causal role in behavior

Face patches have constituted a battlefield for addressing one of the oldest questions in neuroscience: whether discrete regions of the brain are specialized for discrete tasks. Many of the physiological properties of face cells discussed so far suggest that they subserve a particular behavioral function, namely, enabling a monkey to detect faces and recognize their identity. For example, the role for face patches in face detection is supported by the match between the contrast template coded by ML cells and the statistical properties of faces in the real world illuminated from different directions (Fig. II-3d). Similarly, the role of face patch cells in identification is supported by the finding that facial identity can be linearly reconstructed from small numbers of these cells (Fig. II-5d). Yet, these findings do not rule out the possibility that the actual cells responsible for face detection and identification are located elsewhere in the brain, interspersed with cells responsible for other tasks. Here, we first review behavioral evidence that monkeys are adept at detecting and recognizing faces, and then review evidence for the causal role of face patches in these functions.

Monkeys are able to detect faces starting in infancy, with baby macaques tending to look at pictures of humans and monkeys over non-face objects (Sugita, 2008) (this tendency requires prior face experience though, see Box 3, “Development”). They share this face detection ability with human infants who, from birth, preferentially track faces over non-face objects (Goren et al., 1975; Johnson et al., 1991; Valenza et al., 1996). As adults, humans are experts at face recognition, capable of memorizing large numbers of faces over many years (Bahrick et al., 1975) and recognizing faces more accurately and quickly than other visual stimuli (Yin, 1969). Macaques are also adept at individual face recognition. A 1979 study by Rosenfeld and Van Hoesen found that rhesus macaques are able to discriminate faces of conspecifics invariant to changes in orientation, posture, size, color, or illumination (Rosenfeld and Van Hoesen, 1979). Several studies since confirmed that monkeys can recognize faces (Bruce, 1982; Overman Jr and Doty, 1982; Wright and Roberts, 1996). More recently, Moeller et al. found that monkeys could readily learn to discriminate a large number of new human identities, showing almost immediate generalization from delayed match to sample on a set of five faces to a new set of 32 different faces (Moeller et al., 2017); the recognition demand in this study was especially high: the 32 faces were all synthetic, hairless, and identical in complexion, and expressions of the same individual were varied to preclude matching based solely on low-level image features. Studies exploring the spontaneous, untrained viewing behavior of macaque monkeys also support a natural ability of macaques to recognize familiar faces. For example, Gothard et al. presented a pair of faces of the same monkey during a familiarization trial, followed by a pair of faces consisting of the just seen monkey (with altered gaze/view) and a new monkey (Gothard et al., 2009). Monkeys strongly preferred to look at the novel face, indicating that they could distinguish the two faces (see also (Pascalis and Bachevalier, 1998)).

Despite this evidence for strong face recognition capabilities in macaques, a recent review has argued that monkeys are very poor at individuating faces of conspecifics, unlike humans, and hence the macaque face patch system is not relevant for understanding the neural basis for human face recognition (Rossion and Taubert, 2019). The authors cited papers showing that monkeys need many trials to discriminate a small number of faces, show poor performance even after long training, and generalize poorly to novel images of the same individuals or images of new individuals (e.g., (Parr et al., 2008)). We believe the discrepancies in the monkey face recognition behavioral literature stem from the fact that monkeys cannot be instructed verbally, and hence generalization strategies as well as overall motivation of monkeys may differ widely across different labs and different training paradigms. Given the existence of studies clearly showing strong face recognition ability in macaques, we believe the poor face recognition performance observed in some behavioral studies is not due to a fundamental inability of monkeys to recognize faces proficiently, but rather to specifics of training procedures in those studies (e.g., use of food vs. water reward, use of a free viewing touch screen paradigm vs. a fixation paradigm, etc.). Next, we review causal intervention studies probing the specific role of the face patch system in face detection and recognition.

An early study aimed to test the necessity of cortex in the upper and lower banks of the STS for face perception. Heywood and Cowey bilaterally removed both banks and the fundus of the STS, extending from AP +5 mm to +25 mm (Heywood and Cowey, 1992). Prior to the lesion, the animals learned various face-related tasks: to discriminate between faces and inanimate objects, to select the odd face from a group, to inspect a face then select the matching face from a pair of faces, to discriminate between novel and familiar faces, and to discriminate between direct versus averted gaze in faces. Surprisingly, the only task in which animals were impaired following STS removal was gaze discrimination. The authors concluded that the “face-cell area of the brain” is primarily concerned with representation of social signals and not facial identification. This conclusion is clearly at odds with recordings from IT face patches showing a strong signal related to representation of invariant facial identity. The discrepancy can be readily explained: the six IT face patches are not confined to the upper and lower banks of the STS. While patches ML, MF, and AF are in the STS (with ML on the outer lip), patch PL is on the lateral surface, patch AL is typically a few mm ventral of the outer lip on the lateral surface (though sometimes it is directly on the outer lip), and patch AM is on the ventral surface of the brain, far from the STS. Thus it is possible that PL, AL, and AM were all spared in the lesion, and activity in these three patches alone is sufficient to support performance of the various face-related tasks tested by Heywood and Cowey.

To explore the causal role of face cells in face detection, Afraz et al. electrically stimulated face-selective clusters in IT in monkeys trained to judge whether noisy visual images depicted faces or not (Afraz et al., 2006). Microstimulation of face-selective sites biased animals to detect faces, and the magnitude of the effect correlated with the degree of face selectivity of the site. This result suggests a causal relationship between activity of face cells and face detection.

To explore the causal role of face cells in face identification, Afraz et al. trained animals to perform a face gender discrimination task and then applied optogenetics to inhibit cells in a face-selective cluster of macaque IT cortex (Afraz et al., 2015). Silencing a face cluster decreased performance by 2%, while silencing an adjacent region had no significant effect. Due to the tiny effect size, the question of whether face patches truly have a privileged role in representing faces remained unanswered. Furthermore, the study did not specifically target fMRI-defined face patches.

In a microstimulation study, Moeller et al. specifically stimulated face patches while macaques performed a delayed match-to-sample task (Moeller et al., 2017). An earlier electrical microstimulation study in humans found that stimulation of the FFA caused human subjects to perceive dramatic distortions of the experimenter's face ("You just turned into somebody else, your face metamorphosed; your nose got saggy and went to the left; you almost looked like somebody I'd seen before, but somebody different") (Parvizi et al., 2012; Schalk et al., 2017). However, because human subjects are available for only a short period of time, parametric exploration of microstimulation effects in humans is challenging. Consistent with the human findings, Moeller et al. found that stimulation of macaque face patches had a large effect on the percept of face identity. When monkeys saw two identical faces, they reported them as identical ~90% of the time; but when a face patch was electrically stimulated during presentation of the second face, judgement of "same identity" went down to ~10% (Moeller et al., 2017). The effect depended strongly on precise targeting to the center of a face patch, underscoring the privileged role of face patches in coding facial identity. The same effect was found for stimulation of multiple face patches (ML, MF, AL, AF, and AM), supporting the idea that these patches work in concert to generate face perception.

Does microstimulating a face patch influence the perception of non-face objects? Indeed, given that people can see faces in cars, bread, buildings, and electrical outlets, what is a face for a face patch? To address the specificity of face patches for representing real faces, Moeller et al. microstimulated face patches while monkeys observed a variety of non-face objects. They found that microstimulation of face patches does not affect the percept of clearly non-face objects. Surprisingly, face patch stimulation does have a significant effect on the percept of face-compatible non-face objects, including apples, citrus fruits, and cartoon houses. Thus it is possible that face patches are used by the brain not only to represent faces, but also non-face objects eliciting weak but significant responses (Haxby et al., 2001) (conversely, these responses may be why we can see faces in non-face objects).

A study by Sadagopan et al. inactivated face patch ML pharmacologically using the GABA agonist muscimol while monkeys performed an object detection task (Sadagopan et al., 2017). Monkeys were trained to find an object from one of three categories (human faces, macaque bodies, shoes) embedded in a cluttered scene on a touch screen and select it by pressing the object. Inactivation of face patch ML reduced face detection performance by 11%, whereas detection of the other categories was unimpaired. Thus, even inactivation of only one of the 12 IT face patches causes a change in behavior specific to faces. When a nearby region outside the face patches was pharmacologically inactivated, face detection was not significantly reduced.

In sum, evidence is clear that face patches play an important causal role in face detection and recognition, and this effect is demonstrably absent for neighboring IT regions outside face patches. Zooming out, the space of face-related behaviors that have been tested during face patch recordings and stimulation so far has been extremely limited. Little is known about the causal role of face patches in judgements of expression, gaze, head orientation, and other changeable features. Furthermore, the role of face patches during visual search, episodic recall, and other naturalistic behaviors remains unknown.

Interaction with other brain areas

One cannot learn the rules of chess by observing only one chess piece per game. The face patches are not a closed system. They provide essential input to downstream regions responsible for orchestrating multiple face-dependent behaviors. For example, to interpret the state of mind of other agents or judge the social context of a situation, multiple aspects of a face need to be read out including identity, expression, gender, age, ethnicity, attractiveness, and familiarity.

Simultaneously observing activity in face patches and connected functional modules offers a paradigm to understand how two brain regions interact to fulfill their respective purposes. In this section, we first review what is known about interactions between face patches before moving to interactions between face patches and downstream brain areas associated with diverse functions including memory, emotion, and value computation.

Interactions between IT face patches*.* The computations performed by the face patch system are not a simple feedforward hierarchy from posterior to anterior, culminating in a view-invariant representation of facial identity in face patch AM. Tracer and microstimulation experiments show that the face patches are densely and recurrently connected rather than a unidirectional chain of modules (Moeller et al., 2008; Grimaldi et al., 2016). This suggests that recurrent computations occur, where the signal takes several iterations through the face patches. The recurrent connectivity of the face patch system carries potential to perform more powerful computations than the same network with only feedforward connections. One possibility is that the feedback connections are used for learning or to deploy attention. Another possibility is that for challenging visual scenes that cannot be recognized at a glance in a single feedforward sweep, the recurrent computations help resolve ambiguities (Tang et al., 2018a; Kar et al., 2019).

An observation by Ohayon et al. provides an entry point to this problem (Ohayon et al., 2012). They found that in response to an inverted contrast face, face cells in ML/MF show an increase in response latency of about 50 ms, while their response magnitude stays the same. Presumably the contrast-inverted features do not match the feedforward filters of cells in the middle face patch, so the initial response of these cells, which occurs around 100 ms after face onset, is omitted. What signal triggers the delayed response? Does AM or another higher-level structure recognize the inverted contrast face as a face, or are multiple recurrent iterations within the face patches necessary to resolve the contrast-inversion? Currently, we still have very little understanding of detailed network dynamics across the face patch system in situations like this, where the face input is degraded. Interestingly, if the contrast-inverted image is cropped so that the outline and hair of the face are not visible, the response becomes not only delayed but also strongly reduced in amplitude (Ohayon et al., 2012), suggesting that even recurrence cannot fully restore responses in this situation.

Schwiedrzik and Freiwald (Schwiedrzik and Freiwald, 2017) found evidence that feedback from higher- to lower-level face patches may implement predictive coding (Rao and Ballard, 1999; Friston, 2009; Huang and Rao, 2011): While monkeys passively learned sequences of faces, they observed prediction error signals in face patch ML whenever an unexpected face was shown during the sequence. However, unlike the face tuning in ML, the tuning of this prediction error was view-invariant and identity-specific, as in anterior face patches. The authors argued that the prediction error in ML may thus be computed from a top-down prediction from more anterior face patches.

Interactions with claustrum and consciousness. Both tracer and stimulation studies reveal strong connections between face patches and the ventral part of the claustrum (Moeller et al., 2008; Grimaldi et al., 2016). The claustrum is a thin sheet of gray matter adjacent to cortex that is remarkable in that it connects reciprocally to the entire cortex (Pearson et al., 1982; Tanné-Gariépy et al., 2002; Fernández-Miranda et al., 2008). Due to its broad connectivity, it has been speculated to be involved in cross-modal integration (Sherk, 1986; Ettlinger and Wilson, 1990; Edelstein and Denaro, 2004) and binding of features into a unifying conscious experience (Crick and Koch, 2005), although its precise function remains unclear. Thus inputs from the face patches may be broadcast via the claustrum throughout the brain and inputs from the claustrum to the face patches may provide contextual information from other brain areas. This way, the claustrum may serve as a hub between areas that are not directly connected with the face patches but still require face-related information or provide important information to the face patches.

Related to the question of whether the claustrum is the seat for consciousness (Crick and Koch, 2005) is the question of whether face patches themselves reflect the conscious percept of a face or just measure the physical visual input. Both the human FFA and 90% of monkey IT cells respond to reported switches of the conscious percept in binocular rivalry (Tong et al., 1998). It has been debated whether these modulations reflect the conscious percept or just the cognitive factors associated with actively reporting the percept (Frässle et al., 2014; Tsuchiya et al., 2015; Overgaard and Fazekas, 2016). Recently, however, electrophysiological recordings targeted to face patches found that in both ML and AM, a majority of cells follow perception even in a binocular rivalry paradigm without report (Hesse et al., 2019).

Interactions with pulvinar and attention*.* The pulvinar has been implicated in selective attention (Desimone et al., 1990; Robinson and Petersen, 1992; Olshausen et al., 1993; Shipp, 2004; Saalmann and Kastner, 2011), as deactivation leads to attentional impairment (Petersen et al., 1987; Desimone et al., 1990; Wilke et al., 2010; Wilke et al., 2013). As a possible mechanism, it has been proposed that pulvinar induces synchrony between IT and retinotopic regions to preferentially route behaviorally relevant information between the two areas (Saalmann et al., 2012). The pulvinar may be especially important when paying attention to objects rather than just locations in visual space, as it is preferentially activated when subjects pay attention to a region bound by an object compared to an unbound region of space (Arrington et al., 2000).

The purpose of the reported connections between face patches and pulvinar may thus be twofold. First, faces strongly attract spatial attention, demonstrated by the fact that when free-viewing images, subjects rapidly saccade to faces about 16 times more often than to similar regions containing no faces (Cerf et al., 2009). The face patch → pulvinar connection may thus entail a mechanism by which detection of a face automatically draws spatial attention to that region. Second, during active search for a face the reciprocal pulvinar → face patch connection may mediate feature-based attention to faces, facilitating their detection in clutter (Maunsell and Treue, 2006). Recording from face patches and pulvinar simultaneously while monkeys perform various attention tasks may shed more light on these mechanisms.

Interactions with amygdala and emotion processing. The anterior face patches form connections with the basal nucleus, accessory basal nucleus, and lateral nucleus of the amygdala (Moeller et al., 2008; Grimaldi et al., 2016). The amygdala contains large proportions of cells that respond selectively to faces (Sanghera et al., 1979; Nakamura et al., 1992; Fried et al., 1997; Calder and Nummenmaa, 2007; Kuraoka and Nakamura, 2007; Rutishauser et al., 2011). Many of these neurons encode the perceived expression of the face (Wang et al., 2014), and damage to the amygdala causes impairments in recognizing expressions but not identity (Adolphs et al., 1994). Amygdala neurons in both humans and monkeys respond more strongly to faces of their own species (Sigala et al., 2011; Minxha et al., 2017). Moreover, amygdala cells respond when human subjects saccade to a face (Minxha et al., 2017) or when monkeys make eye contact (Mosher et al., 2014). Lesioning monkey amygdala eliminates their preference to look at faces (Taubert et al., 2018), suggesting that the amygdala may also play a role in establishing the salience or value associated with faces.

Compared to other objects, faces are a unique in how much information about the emotion and intention of another agent can be inferred from them. It remains an open question if and how face patches and amygdala work together to achieve this feat. One possibility is that the amygdala builds a code for expression and gaze using information about the physical properties of a face relayed from the face patches. Alternatively, the amygdala may only receive information about facial identity from IT face patches, and acquire expression/gaze information through an independent pathway.

Interactions with medial temporal lobe and memory. Grimaldi et al. found connections between face patches and perirhinal cortex as well as parahippocampal regions TF/TH and TFO (Grimaldi et al., 2016). These regions have been implicated to play a role in memory, recency, and familiarity (Fahy et al., 1993). It is possible that the connections between face patches and the medial temporal lobe help to recognize a face as familiar and evoke all the memories of the individual associated with it. Consistent with this, face patch PR in perirhinal cortex is modulated by the familiarity of faces (Landi and Freiwald, 2017). However, it remains unclear whether this familiarity coding is inherited from face patches in IT or if it arises initially in PR or an ever more downstream area.

The interaction between face patches and the medial temporal lobe may also may provide a tractable entry point to studying imagination (Kornblith and Tsao, 2017). When a subject is asked to visualize the face of a person, the human FFA is activated (O'Craven and Kanwisher, 2000). Is the code used by face patches to represent an imagined face the same as the code used to represent a physically presented face? If so, how is the memory of a familiar face transformed into the perceptual representation of a face during visual imagery?

Interactions with frontal cortex and value-based decision making. Tracer studies have found only weak connections between anterior face patches and orbitofrontal area 13 as well as ventrolateral prefrontal cortex (Grimaldi et al., 2016). However, as mentioned above, a simple face localizer scan reveals three face-selective patches in prefrontal cortex, within ventrolateral, orbifrontal, and pre-arcuate regions (Moeller et al., 2008). These regions are activated by electrical microstimulation of IT face patches (Moeller et al., 2008), suggesting they may inherit their selectivity through interactions with IT face patches. It is believed that orbitofrontal cortex is used for value judgement and lateral prefrontal cortex then uses these values to compute decisions on which actions to take (Wallis, 2007). Faces can have very different values depending on their gender, age, expression, and face cells in both orbitofrontal (Barat et al., 2018) and ventrolateral prefrontal cortex (Romanski and Diehl, 2011) show selectivity for these social dimensions. This opens an avenue for studying how value is computed from identity information encoded in IT face patches and how prefrontal cortex then uses this information to inform decisions and behavior.

Conclusion and boxes

Returning to Marr's question, whether finding a face cell would tell us anything much at all, we think the answer is a resounding yes. The clustering of face cells into distinct patches performing different steps of face processing has made it possible to identify the sequence of transformations in face representation and decipher the code for representation of facial identity used by these patches. This knowledge now gives us a powerful key to unlocking not only object recognition in general, but also higher-level functions including memory, attention, consciousness, and decision making that operate on faces. When trying to decipher an encrypted communication, the key step is to identify the meaning of the first word. Using the same approach of fMRI and targeted electrophysiology that has been instrumental to dissecting the face patch system, several additional networks in IT have now been discovered, including color patches (Lafer-Sousa and Conway, 2013; Chang et al., 2017), scene patches (Kornblith et al., 2013), and body patches (Tsao et al., 2003a; Popivanov et al., 2012; Kumar et al., 2017). Most strikingly, the recent discovery of a set of parallel IT networks tiling object space, each harboring the same anatomical organization and object coding mechanism as the face patch network, suggests that the face patch network is truly a model for all of IT cortex (Bao et al., 2019). For understanding higher-level cognition beyond IT, our knowledge of the code for facial identity in AM makes it possible to transcend studying shadows of cognitive processes, to directly study the link between a palpable high-level percept and the downstream cognitive processes it triggers. The turtle’s underbelly is staring us in the face.

Boxes

Box 1: Are faces special?

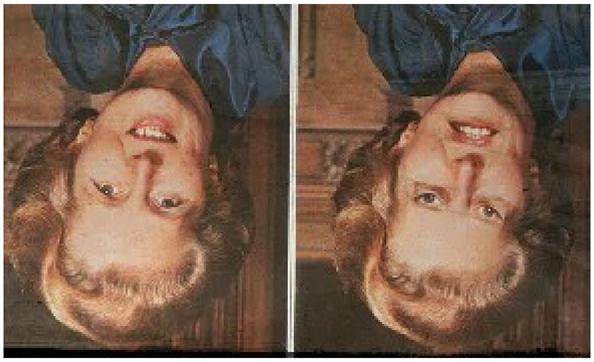


Figure II-B1: The Thatcher illusion. *At first glance, neither the left nor the right image seem particularly odd since the faces are upside-down. However, if one turns the page upside-down, one notices that the local features in one of the faces have been inverted. This demonstrates the impairments of face recognition when faces are presented upside-down, and simulates the effect of a face patch lesion. Adapted from (Thompson, 1980).*

Are faces special, in the sense that special neural mechanisms are used to process faces that are not used for other object classes? And if so, do monkeys also show hallmarks of this special form of processing? Evidence that humans are experts at recognizing faces comes from several behavioral effects including the composite effect (Young et al., 2013) (combining the top and bottom halves of two different faces by aligning them interferes with the percept and identification of the constituents), the whole-part effect (Tanaka and Farah, 1993) (a face feature is more easily recognized if it is part of a whole face rather than in isolation), and the face inversion effect (Yin, 1969; Valentine, 1988) (a face is more easily recognized when upright than when upside down). In monkeys, the face inversion effect has been controversial, with some studies claiming that monkeys do show recognition differences between upright and inverted faces (Parr et al., 1999; Gothard et al., 2009), and others claiming that they do not (Rosenfeld and Van Hoesen, 1979; Bruce, 1982; Parr et al., 2008). Freiwald et al. found that cells in ML become less tuned to facial features of cartoon faces when the faces are inverted (Freiwald et al., 2009). Taubert et al. reported that face-selective cells in ML and AL but not outside of face patches preferentially encode upright faces compared to inverted faces (Taubert et al., 2014). This finding suggests functional differences between face-selective cells inside face patches vs. those scattered across IT outside of face patches, and the possibility that specialized processing for faces may be restricted to the former. An effect related to the face inversion effect is the Thatcher illusion (Thompson, 1980): when a face is presented upside-down and local features such as the mouth and the eyes are relatively inverted, the face does not appear unusual (see Fig. II-B1); however, if the whole face is reverted back to upright, the inverted local face features become strikingly apparent. Monkeys show this configural effect behaviorally (Adachi et al., 2009). Taubert et al. found a neural correlate of the Thatcher illusion in face patch ML but not in AL or IT cortex outside of face patches (Taubert et al., 2015) (see also (Sugase-Miyamoto et al., 2014)).

Tan and Poggio argued that the three behavioral markers of specialized holistic face processing (composite effect, whole-part effect, and inversion effect) can all be modeled using the single neural factor of neural tuning size (Tan and Poggio, 2016), i.e., there is nothing fundamentally different about face processing compared to general object processing. In their HMAX model (Riesenhuber and Poggio, 1999), which contains two layers each consisting of simple and complex cells, the neural tuning size is defined as the size of the template of simple cells in the second layer, which detect parts or features of the face. For small neural tuning sizes, where the template covers only one face part such as the eye or nose, the output of the model is non-holistic and more like general object recognition, whereas for larger neural tuning sizes, where the template covers multiple face parts, the output of the model is holistic like face recognition and reproduces the three behavioral markers described above.

A theoretical study by Leibo et al. offers a computational perspective on why face recognition should be separated from recognition of other objects in the brain (Leibo et al., 2011). They argue that the transformation of a face image evoked by 3D rotation of the face is class-specific and different from the transformations of other objects. By training a class of models on 3D rotations of faces, they show that the models learn to generalize to other faces and can extract their identity in a view-invariant way but fail on other object classes. The argument that identity-preserving image transformations are class-specific thus suggests the need for specialized modules for specific classes such as faces.

These computational studies suggest that processing of faces is not that unique after all: classic markers of a supposedly unique "holistic" processing style can be boiled down to increased receptive field size of face cells, and modularity can be explained by a principle, class compatibility of 3D transformations, that applies to all objects. The finding that at least half of IT cortex shares similar anatomical organization and coding principles as the face patches (Fig. II-6) further supports the idea that processing of faces is not unique.

Box 2: Face patches across species

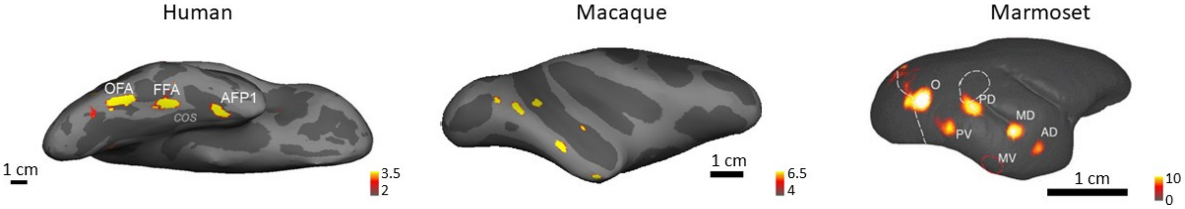


Figure* II*-B2: Face patches across different primate species. *Face patches mapped in the human, macaque, and marmoset are shown. Human and macaque maps adapted from (Tsao et al., 2008a), marmoset map adapted from (Hung et al., 2015).*

Face-selective areas have been identified using fMRI in humans (Kanwisher et al., 1997), macaques (Tsao et al., 2003a), and marmosets (Hung et al., 2015) (Fig. II-B2). In addition, there is a report of a face area in dogs (Cuaya et al., 2016). Face cells have been reported not only in humans (Khuvis et al., 2017) and macaque monkeys (Gross et al., 1972), but also in sheep (Kendrick and Baldwin, 1987). Finally, the behavioral ability to identify faces has been reported in a wide range of organisms including cattle (Coulon et al., 2009), pigeons (Stephan et al., 2012), archer fish (Newport et al., 2016), wasps (Tibbetts, 2002), and crayfish (Van der Velden et al., 2008).

The study of face processing is a wonderful example of the fruitful interplay between human and macaque research and underscores the importance of non-human primate research. The discovery of face areas in humans was one of the major motivations to look for such areas in monkeys. In turn, understanding of face areas in monkeys has shed new light on face areas in humans, and even instigated the discovery of a new face area in the human anterior temporal lobe (Tsao et al., 2008b; Rajimehr et al., 2009). Monkey electrophysiology experiments also reveal the limitations of fMRI and the importance of recording from single cells to understand what an area is coding: Dubois et al., using MVPA on fMRI data from face patch AM, were unable to decode any identity information (Dubois et al., 2015), whereas single-unit reconstructions of face identity from AM activity are strikingly accurate (Fig. II-5d). Because brain structures supporting face processing exist across primate species, experiments can be designed that maximally leverage the advantage of each species: deep probing of neuronal mechanisms in non-human primates, and use of complex task designs exploiting language and introspection in humans.

What is the homology between human and macaque face areas (Tsao et al., 2008a; Yovel and Freiwald, 2013)? Human face areas include the occipital face area (OFA), fusiform face area (FFA), several anterior temporal face areas, and a face area in the superior temporal sulcus (STS-FA). Evidence has been put forward that these areas have different functional specializations, with the OFA involved in processing face parts (Pitcher et al., 2011), the STS-FA in processing changeable aspects such as gaze direction (Haxby et al., 2000; Pitcher et al., 2011), and the FFA in processing identity (Rotshtein et al., 2005). One might guess based on spatial position and functional properties that human OFA, FFA, STS-FA, and anterior face patches correspond respectively to macaque PL, ML/MF, upper bank MD, and AL/AF/AM. However, fMRI-guided single-unit recordings in humans will be necessary to make definitive conclusions concerning homology (Khuvis et al., 2018).

The discovery of face areas in marmosets is an exciting recent development that portends even deeper understanding of the development and function of face patches (Hung et al., 2015), given the short generation time of marmosets and the wide range of tools available for marmoset neural circuit dissection, including large-scale surface electrode arrays, two-photon imaging, and CRISPR gene editing. So far, ECOG recordings from the marmoset confirm existence of multiple face patches (Hung et al., 2015), but the detailed single-unit properties of individual fMRI-identified marmoset face patches remains unknown.

Box 3: Development of face patches

How do face patches develop? Are the locations of face patches already laid out during development by the primate's genetic program or does their development require experience with faces? In humans, Dehaene et al. found that literacy induces increased activation to text in a brain region specialized for representing word forms from a familiar language called the "visual word form area." This occurred regardless whether literacy was acquired during childhood or adulthood (Dehaene et al., 2010). The increased response to text comes with the cost of a slightly smaller face-selective area compared to illiterates. On the other hand, Deen et al. found that in human infants a few months after birth, extrastriate cortex shows modules that respond preferentially to faces or scenes with a spatial organization similar to adults, which are then refined throughout development (Deen et al., 2017). In monkeys, some functional organization of face-selective regions exists already one month after birth, which is then refined and stabilized into the face patch system within two years (Arcaro and Livingstone, 2017). Importantly, experience with faces appears to be necessary for these modules to acquire their selectivity: Arcaro et al. raised monkeys without exposure to faces, and these monkeys did not show any domains selective for faces (Arcaro and Livingstone, 2017). Moreover, the face-deprived monkeys did not show preference for looking at faces over other objects.

These findings suggest the importance of experience in IT development. Nevertheless, it is possible that IT already possesses a proto-organization into modules at birth that is then activated by experience with a specific object class. Indeed, IT cortex is organized retinotopically, albeit much more coarsely than retinotopic cortex (Janssens et al., 2014; Kolster et al., 2014), with face patches lying in the foveal zone (Janssens et al., 2014). Unlike selectivity to object categories, this retinotopic organization is already present at birth (Arcaro and Livingstone, 2017), i.e., before visual experience. Srihasam et al. trained juvenile monkeys to recognize three sets of artificial shapes (Srihasam et al., 2014). The three sets consisted of “Helvetica” symbols, Tetris shapes, and cartoon face symbols, respectively. Following intense exposure to these shapes, monkeys developed patches in IT for each set that were selectively activated by shapes in that set. Even though the order in which monkeys were introduced to the three sets of shapes was different for each monkey, the locations of the patches for each set was similar across monkeys. Together, these results suggest that IT has a retinotopic proto-organization inherited from inputs from earlier visual areas (Hasson et al., 2003), and exposure to a certain class of objects causes the formation of category-selective domains at specific locations within this retinotopic map. These retinotopic locations may be influenced by where the monkey tends to look (e.g., faces are represented in the foveal zone because monkeys tend to foveate them) as well as intrinsic properties of an object class (e.g., curvature) that are preferred by locations with a certain eccentricity (Levy et al., 2001; Srihasam et al., 2014). Besides coarse retinotopy, topography in the IT map is governed by clustering in object space (Fig. II-6). This additional clustering could be enforced by purely self-organizing principles independent of both genetic hard-wiring and retinotopic selectivity of inputs (Erickson et al., 2000).

Chapter III

FIGURE-GROUND SEGMENTATION

Consistency of border-ownership cells

From: Hesse and Tsao (2016), "Consistency of border-ownership cells across artificial stimuli, natural stimuli, and stimuli with ambiguous contours." In: *Journal of Neuroscience* 36.44 (2016): 11338-11349.

Abstract. Segmentation and recognition of objects in a visual scene are two problems that are hard to solve separately from each other. When segmenting an ambiguous scene, it is helpful to already know the present objects and their shapes. However, for recognizing an object in clutter, one would like to consider its isolated segment alone to avoid confounds from features of other objects. Border-ownership cells (Zhou et al., 2000) appear to play an important role in segmentation, as they signal the side-of-figure of artificial stimuli. The present work explores the role of border-ownership cells in dorsal macaque visual areas V2 and V3 in the segmentation of natural object stimuli and locally ambiguous stimuli. We report two major results. First, compared to previous estimates, we found a smaller percentage of cells that were consistent across artificial stimuli used previously. Second, we found that the average response of those neurons that did respond consistently to the side-of-figure of artificial stimuli also consistently signaled, as a population, the side-of-figure for borders of single faces, occluding faces and, with higher latencies, even stimuli with illusory contours such as Mooney faces and natural faces completely missing local edge information. In contrast, the local edge or the outlines of the face alone could not always evoke a significant border-ownership signal. Our results underscore that border ownership is coded by a population of cells, and indicate that these cells integrate a variety of

cues, including low-level features and global object context, to compute the segmentation of the scene.

Significance. In order to distinguish different objects in a natural scene, the brain must segment the image into regions corresponding to objects. The so-called “border-ownership” cells appear to be dedicated to this task, as they signal for a given edge on which side the object is that owns it. Here, we report that individual border-ownership cells are unreliable when tested across a battery of artificial stimuli used previously but can signal border-ownership consistently as a population. We show that these border-ownership population signals are also suited for signaling border-ownership for natural objects and at longer latency, even for stimuli without local edge information. Our results suggest that border-ownership cells integrate both local, low-level and global, high-level cues to segment the scene.

Introduction. The two most important tasks of the ventral stream of visual cortex are arguably segmentation of the visual scene and object recognition. Segmentation tells us which groups of pixels in a scene constitute the fundamental units that we can interact with, and recognition gives these units a meaning by telling us what they are. Often, vision is considered as a sequence of processing steps in a feedforward hierarchy (Marr, 1982), where recognition of objects happens after a series of nonlinear operations on the input image (Riesenhuber and Poggio, 2002). Since segmentation, which is thought to happen in retinotopic cortex, is earlier in the feedforward hierarchy of visual areas, it is often assumed to be a necessary step to be completed before recognition, so that the segmented regions corresponding to the object surfaces can be fed to inferotemporal cortex (IT) and recognized individually (Rubin, 1958; Nakayama et al., 1995; Driver and Baylis, 1996). On the other hand, psychophysical studies (Peterson and Gibson, 1993, 1994; Peterson and Kim, 2001; Grill-Spector and Kanwisher, 2005) have suggested that object recognition influences or even precedes segmentation. For example, Peterson and Gibson (1993) found that recognition of object shape can overwrite depth order defined by disparity. Moreover, observers asked to report their first perceived figure-ground organization are influenced by symmetry and orientation-dependent object recognition processes and are more likely to perceive regions as figure compared to ground if they match their object memory (Peterson and Kim, 2001)

One of the most insightful neurophysiological findings for understanding how the brain segments the visual scene are the remarkable border-ownership cells discovered by von der Heydt and colleagues (Zhou et al., 2000). Border-ownership cells are thought to be crucial for segmentation, as their responses signal the side-of-figure for a number of artificial stimuli. However, it is not entirely clear if and how the side-of-figure signal observed in border-ownership cells aids recognition of objects in natural scenes. Conversely, it is not known whether object recognition in IT can influence border-ownership signals in retinotopic cortex.

Like most cells in early visual cortex a border-ownership cell will respond to an edge presented at a given orientation in its receptive field, but it responds differentially depending on the side of the figure that the edge belongs too: a vertical edge can be the border of a foreground object that is either to the left or to the right of it. Zhou et al. showed that for a variety of artificial stimuli, border-ownership cells respond consistently more strongly if the edge belongs to a figure on its preferred side than its non-preferred side, even if the stimuli are locally identical within the receptive field of the cell. Artificial stimuli that were previously shown to evoke consistent border-ownership responses include single luminance squares, occluding luminance rectangles, single and occluding outlines of rectangles, C-shapes (Zhou et al., 2000), disparity-defined squares (Qiu and Von Der Heydt, 2005), and squares evoking the percept of transparent overlay (Qiu and Von Der Heydt, 2007) (see stimuli used in Fig. III-5a-d). Here, we asked whether border-ownership cells can also infer border-ownership for natural objects, with not necessarily straight edges and inhomogeneous, possibly confounding textures. In fact, in natural scenes, segmentation can often be ambiguous based on only local, low-level cues (McDermott, 2004). Consider for example the famous Dalmatian dog display or the camouflaged owl in Fig. III-1. It appears impossible for an algorithm that uses only local, low-level cues to infer the correct segmentation. Yet, once we recognize the Dalmatian dog, we perceive it as an object with a contiguous surface. And we are able to infer the boundary of the owl. Bottom-up, purely feedforward algorithms would likely come to the critically different, erroneous interpretation that it is a texture. Is this perceived segmentation signal for recognizable, natural objects present in the side-of-figure signal of border-ownership cells?

To answer this question, we recorded from border-ownership cells and systematically presented a battery of both artificial stimuli and natural face stimuli, as well as face stimuli with ambiguous contours, in order to find out how cells that respond consistently to the side-of-figure of artificial stimuli would respond to the presented natural object stimuli.

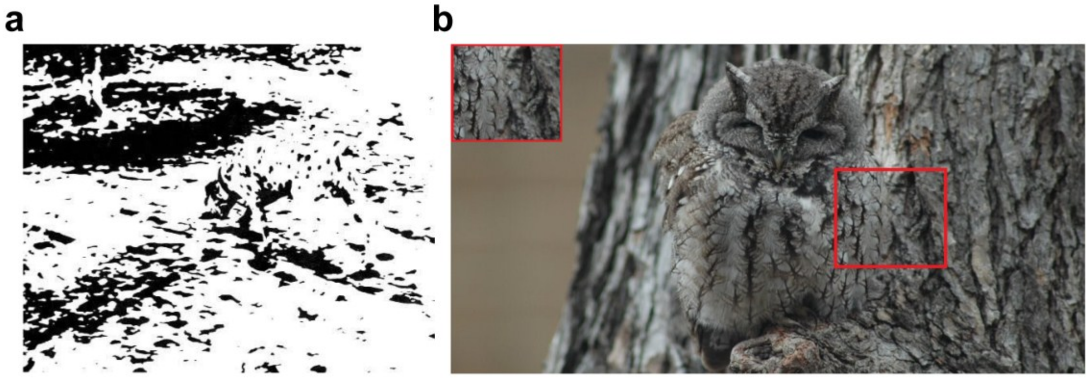


Figure III-3: Segmentation can be ambiguous based on low-level cues. *(**a**) While the famous Dalmatian dog display is considerably more difficult to segment without knowing that there is a dog present, once one recognizes the dog, one also perceives it as a contiguous surface. (**b**) The camouflaged owl seems*

impossible to segment based on local cues (red square, inset), but as one recognizes the owl one can infer and perceive the boundary.

Results. We targeted regions in dorsal V2 and V3 that elicited strong fMRI activation in response to disparity-defined shapes vs. full-field disparity. Fig. III-2 shows responses of a border-ownership cell at a representative location (Fig. III-2a, functional activation overlaid). Fig. III-2b shows the receptive field of the example cell mapped by computing the STA. All analyses below are based on a sample of a total of 201 single units (126 Monkey T, 75 Monkey J) for which we manually verified correct positioning on the receptive field and orientation tuning. This sample is biased, as we were explicitly looking for border-ownership cells and skipped cells that were not promising candidates (see Methods). For Monkey J, where the fMRI signal was weak and we were just guided by anatomical locations where monkey T had shown high functional activation, we found slightly fewer border-ownership cells (p=0.04, two-sided unpaired t-test on average modulation indices across artificial stimuli). Receptive fields were in the lower left and lower right quadrant of the visual field for monkey T and monkey J, respectively and eccentricity ranged from 1° to 5°.

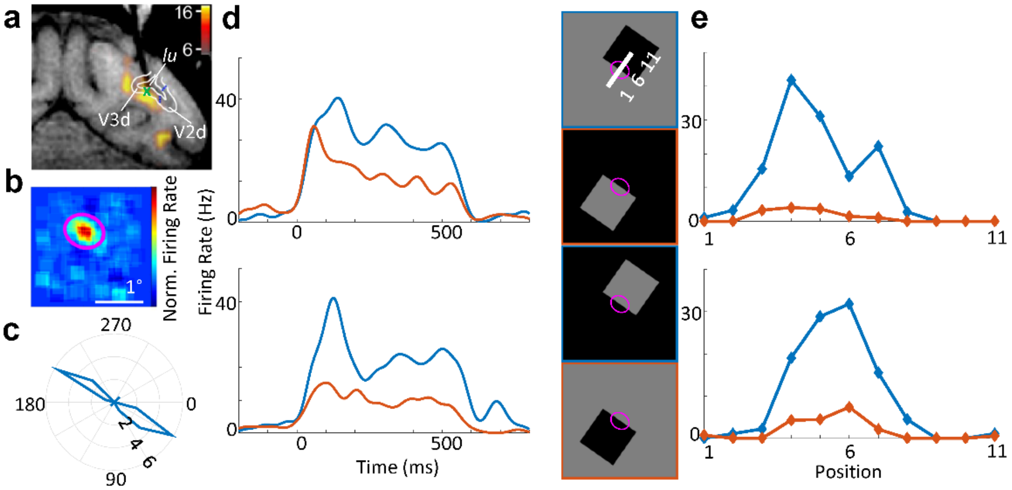


Figure III-2: An example border-ownership cell.

*(**a**) Electrode targeting V3d; fMRI activation for disparity checkerboard vs. full field disparity is overlaid. Green cross: Location of electrode tip. Dotted blue lines: Retinotopically defined area boundaries of V2d and V3d. lu: lunate sulcus. (**b**) Receptive field as computed by the spike-triggered average. (**c**) Orientation tuning of the cell. Radius and angle of the polar plot correspond to firing rate and presented orientation of moving sine grating, respectively. This example cell had a preferred orientation of about 30°. (**d**) PSTHs of responses to luminance squares (presented for 500 ms). Four square stimuli with different contrasts and different sides (right panel) with the edge on the receptive field (purple ellipse). This cell's response was increased when the figure was on the top right side. (**e**) Position test of the cell. To test the robustness of the border-ownership signal across positions within the receptive field, the stimulus was swept across different positions orthogonal to its preferred orientation (x-axis, indicated by the positions 1-11 at the top right of (d)). Across all positions within the receptive field, the response (y-axis) was consistently higher when the figure was on the preferred side of the receptive field. Blue and red conditions are equivalent to the stimuli used in (d). Error bars indicate standard error mean.*

Initially, the primary goal of this study was to determine how border-ownership cells respond to the side-of-figure of natural stimuli. However, preliminary recordings with a variety of artificial stimuli revealed almost no cells that were consistent in their border-ownership preference across all artificial stimuli tested, leading us to carry out a systematic characterization of the consistency of border-ownership cells across a large population of cells and a large battery of artificial stimuli (See Fig. III-5a-d for the battery of artificial stimuli used). Previous studies or border-ownership cells have each focused on specific subsets of artificial stimuli. Thus, a major question remains open: whether there exists a significant fraction of "true" border-ownership cells that signal the side-of-figure reliably across all types of artificial stimuli containing object borders. Moreover, there has been considerable variability in the reported proportions of consistent cells across different studies and stimuli. Zhou et al. (2000) found more than more than half of cells in V2 to be selective to the side-of-figure of single luminance squares; among those cells, 20/42 cells were tuned significantly and consistent to the side-of-figure of occluding squares, 1/42 cells was tuned significantly but inconsistent, and 21/42 cells were not significantly tuned. For C-shapes, Zhou et al. found 4/16 cells to be tuned significantly and consistent, 12/16 to be not significantly tuned, and no cells to be significantly tuned and inconsistent. (Qiu and Von Der Heydt, 2005) found 35% of 174 neurons in V2 to be selective to the side-of-figure of a luminance square, 40% to be selective to depth order, and 21% selective to both, of which 81% were consistent between luminance-defined and disparity-defined side-of-figure. For a transparent overlay stimulus, Qiu and Von Der Heydt (2007) found 127 of 244 of cells to be tuned to the side-of-figure of the luminance square and 30 of those cells to be significantly tuned to the side-of-figure of the transparent bars and consistent with the preferred side-of-figure for the luminance square; they did not report the number of significantly tuned, inconsistent cells.

Paradigm	Significantly consistent		Significantly inconsistent		Not significant	
	All cells	Top 50	All cells	Top 50	All cells	Top 50
Standard test (Single luminance square)	55% (111/201)	84% (42/50)	n/a	n/a	44% (90/201)	16% (8/50)
Occluding squares	19% (21/110)	30% (13/42)	10% (11/110)	2% (1/42)	71% (78/110)	67% (28/42)
Occluding outlines	11% (12/111)	17% (7/42)	16% (18/111)	5% (2/42)	73% (81/111)	79% (33/42)
C-shapes	11% (12/108)	19% (8/42)	22% (24/108)	10% (4/42)	67% (72/108)	71% (30/42)
Transparent	20% (21/107)	26% (11/42)	5% (5/107)	2% (1/42)	76% (81/107)	71% (30/42)
Four squares control	37% (40/107)	55% (23/42)	10% (11/107)	0% (0/42)	52% (56/107)	45% (19/42)
Single full faces	41% (45/111)	69% (29/42)	17% (19/111)	7% (3/42)	42% (47/111)	24% (10/42)

Occluding full faces	14% (16/111)	29% (12/42)	5% (6/111)	2% (1/42)	80% (89/111)	69% (29/42)
Single ambiguous face	20% (22/111)	29% (12/42)	10% (11/111)	0% (0/42)	70% (78/111)	71% (30/42)

Table III-1: Consistency of cells for border-ownership stimuli. *For different stimuli presented during the experiment, we computed the percentage of cells that were significantly tuned to the side-of-figure of the stimulus and consistent with the preferred side-of-figure for the single luminance square. Significance criterion was p<0.01 as in Zhou et al. (2000). For the single luminance square (first row), the proportion of significantly tuned cells merely indicates the proportion among all 201 analyzed cells, and among the 50 cells with highest average modulation index across artificial stimuli, respectively, that were significantly tuned the side-of-figure of the single luminance square by two-way anova on side-of-figure and contrast polarity. For the remaining stimuli, we only considered cells that were significantly tuned to the side-of-figure of the single luminance square, which was the case for 111 out all 201 cells and 42 out of the top 50 cells, and computed the proportions of cells from these two populations that were (a) significantly tuned to the side-of-figure of the stimulus and consistent with the side-of-figure preference for the single luminance square, (b) significantly tuned to the side-of-figure but inconsistent with the single luminance square, or (c) not significantly tuned to the side-of-figure of the given stimulus. Note that 4 of the 111 cells were lost too early to present all artificial stimuli. For comparison, Zhou et al. (2000) found that among the cells that were significantly tuned the side-of-figure of the single luminance square 20/42 cells were tuned significantly and consistent to the side-of-figure of occluding squares, 1/42 cells was tuned significantly but inconsistent and 21/42*

cells were not significantly tuned. For the C-shapes Zhou et al. found 4/16 cells to be tuned significantly and consistent, 12/16 to be not significantly tuned, and no cells to be significantly tuned and inconsistent. For the transparent stimulus, Qiu and Von Der Heydt (2007) found 30/127 cells to be tuned significantly and consistent (note that the significance criterion was p<0.05 in Qiu and Von Der Heydt (2007) and p<0.01 in Zhou et al. (2000)).

Similar to the findings of Zhou et al. (2000), we found that a little more than half of all neurons were significantly selective to the side-of-figure of the single luminance squares (two-way anova p<0.01). Compared to Zhou et al. (2000), however, we found a substantially higher proportion of cells that showed inconsistent border-ownership selectivity when comparing tuning for side-of-figure in luminance-defined figures and another artificial stimulus (Table 1). In particular, for C-shapes and overlapping outlines, we found even more cells that significantly preferred a side inconsistent with the single luminance square preference than cells that preferred a consistent side. For the overlapping luminance squares, the transparent stimulus, and its four square control, in turn, there were more cells that were consistently significantly selective than cells significantly preferring the opposite side, which is more consistent with the findings of Zhou et al. (2000). Demonstrating that we were recording from the same general class of cells as von der Heydt’s group, we reconstructed their population analysis for the transparent stimulus from Qiu and Von Der Heydt (2007) and found qualitatively similar results (Fig. III-3). A subtle difference in the stimuli (Fig. III-3, top) leads to the perception of either two transparent overlaid bars or four rounded squares, which changes the border-ownership at the receptive field location. Although the scale of the marginal distributions of border-ownership modulation indices was somewhat different between our sample of cells and Qiu and Von Der Heydt (2007), we replicated the switching in border-ownership signaled by the cell population.

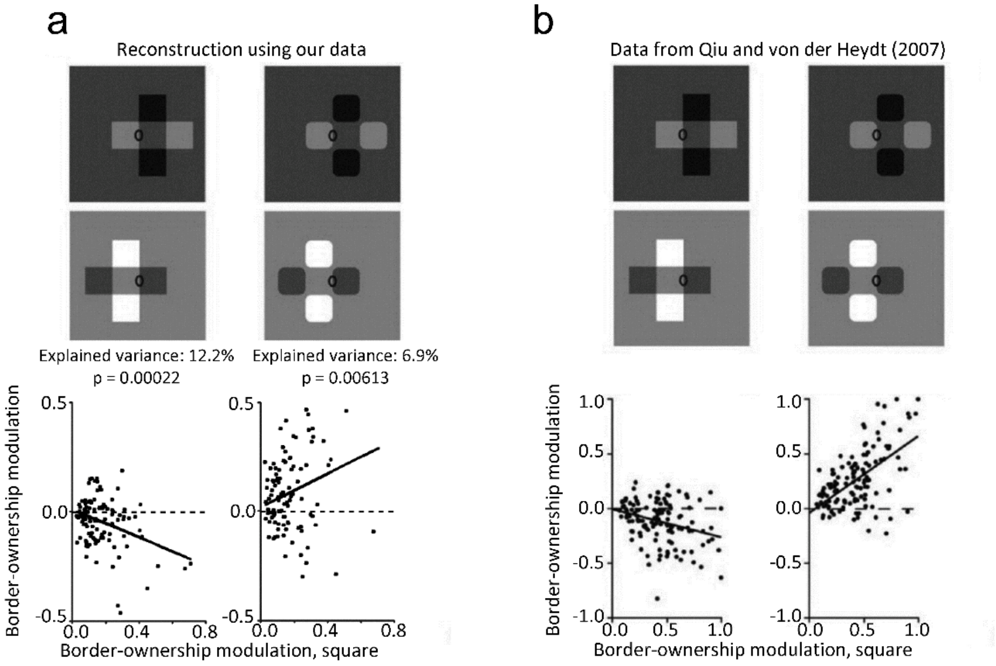


Figure III-3: Comparison of population analysis to Qiu and Von Der Heydt (2007). *We reconstructed the population analysis performed in Qiu and Von Der Heydt (2007) to compare border-ownership tuning of the recorded cells. (**a**) shows the border-ownership modulation index for the single luminance square on the x-axis against border-ownership modulation index of the transparent overlay stimulus (left) and the four square control (right) on the y-axis for all cells we recorded that were significantly tuned to the side-of-figure of the single luminance square. (**b**) shows the original data adapted from Qiu and Von Der Heydt (2007). Note that for this figure, we used the same formula for modulation indices as used in Qiu and Von Der Heydt (2007), which is slightly more complicated than the one in the rest of this article. Thick black lines indicate Standard Model I regression lines, i.e., the ordinary least squares fit of the abscissa to the ordinate. P-values were computed by comparing a transform of the Pearson correlation coefficient to a Student's t-distribution. Note that this represents a relatively conservative measure of correlation. Since the sign of border-ownership for each cell is*

arbitrary, we also computed correlations after duplicating all points by reflecting them over the origin. This measure yielded the same signs of correlations and even higher explained variances ($r^2 = 22.2\%$ *and p* $= 3.15e - 13$ *for transparent overlay stimulus,* $r^2 = 24.7\%$ *and p* $= 9.56e - 15$ *for transparent overlay stimulus).*

We next quantified how many cells were significantly tuned to a given number of artificial stimuli and consistent with the preferred side-of-figure for the single luminance square (Fig. III-4a). We found that 40% of cells were not significantly and consistently tuned for any of the artificial stimuli, and not a single cell was significantly and consistently tuned for all artificial stimuli. There were 11 cells that were tuned consistently across all artificial stimuli, but not significant for all, i.e., the sign of the modulation index, averaged across the two contrast conditions, was the same for each artificial stimulus as for the single luminance square (standard test). In sum, we failed to find a population of cells that were significantly tuned to the side-of-figure across the whole large battery of artificial stimuli in a consistent manner; in particular, many ostensible border-ownership cells that were significantly tuned to the side-of-figure of luminance squares turned out to be inconsistent when tested with a battery of other stimuli. These results strongly suggest that border ownership is encoded by a *population* of cells, which are each, individually, imperfect in their border-ownership selectivity.

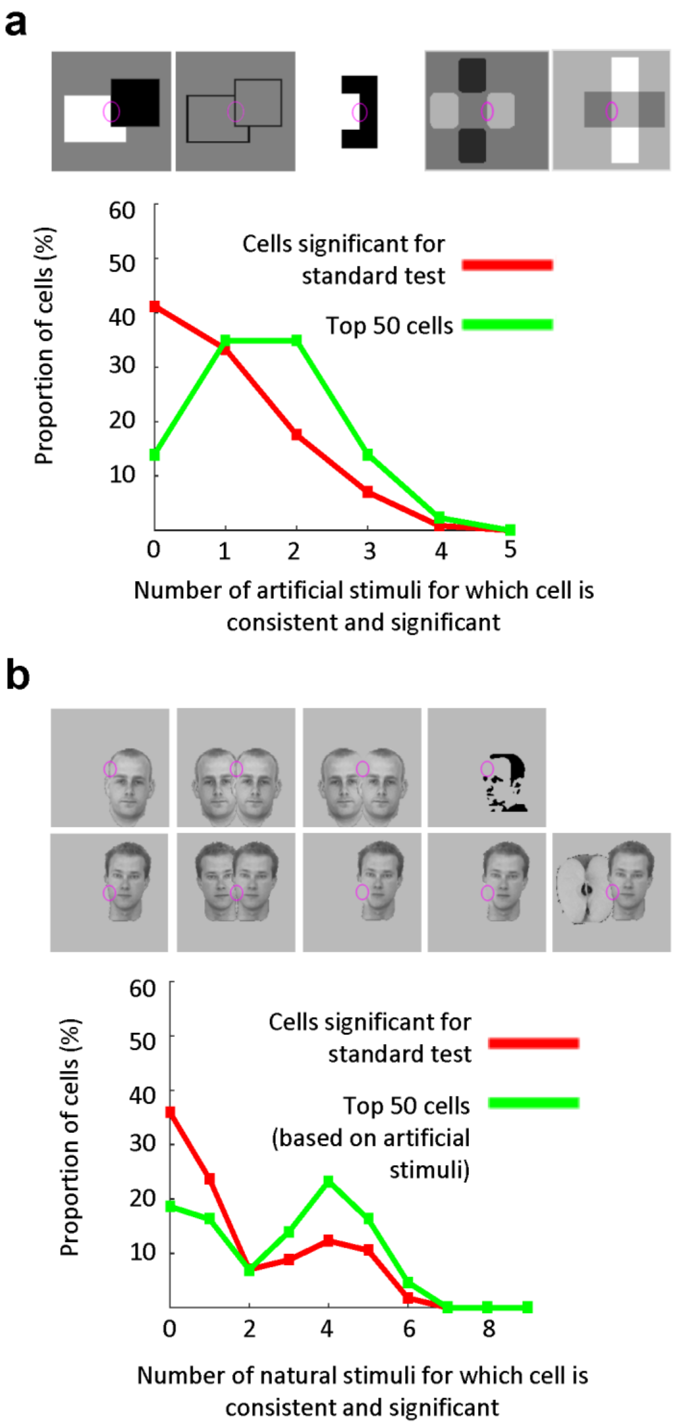


Figure III-4: Quantification of response consistency across artificial and natural stimuli.

*Histograms show the proportions of cells that are consistent and significant for a given number of artificial (**a**) and natural stimuli (**b**). The five artificial stimuli and nine natural stimuli based on which the consistency of the cell was tested are shown at the top of (a) and (b), respectively. A cell is considered consistent for a given stimulus if the modulation index for that stimulus has the same sign as for the single luminance square (standard test). Significance criterion is p<0.01. Red lines show the proportion of cells among all cells that were significantly tuned to the side-of-figure of the single luminance square, whereas blue lines only take into account the 50 cells with the highest modulation index across artificial stimuli, respectively.*

To see whether, despite the inconsistencies of many cells, we could still extract a population signal that is consistent across all artificial stimuli, we next computed the average population response across the most consistent cells. We evaluated each neuron's consistency by computing the average modulation index across all artificial stimuli (shown in Fig. III-5a-d). The average modulation index across artificial stimuli had a mean of 0.038 and a standard deviation of 0.081 across all 201 cells (see also histograms in Fig. III-6, top). We then chose the 50 most consistent cells based on this average modulation index. The exact number of chosen cells did not qualitatively influence the results. Cells were pooled across V2 and V3; when comparing consistency to artificial stimuli between V2d and V3d, we found slightly but significantly more consistent border-ownership signals in V3d than in V2d (p=0.02, unpaired t-test on average modulation indices across artificial stimuli based on 137 cells recorded from V2d and 64 cells recorded from V3d). As can be seen from the green line in Fig. III-4a, by selecting the 50 top cells with this method, the proportion of cells that were consistent across a high number of artificial stimuli substantially increased. Fig. III-5a-d shows the average PSTHs of these 50 selected cells to artificial stimuli. By focusing on the most consistent border-ownership cells, we obtained a border-ownership signal that was reliable across artificial stimuli: the response difference for the standard luminance square test (Fig. III-5a) was strong and consistent, which is expected since we defined the preferred side of each cell based on the luminance square responses. The responses to the other artificial stimuli (Fig. III-5b-d) were also significantly stronger when presented on the preferred side, which is a sanity check that the population of 50 cells indeed represented consistent border-ownership cells. The population border-ownership signal for occluding outlines and C-shapes was weaker than for other artificial stimuli, but was nevertheless consistent (see also Table 1).

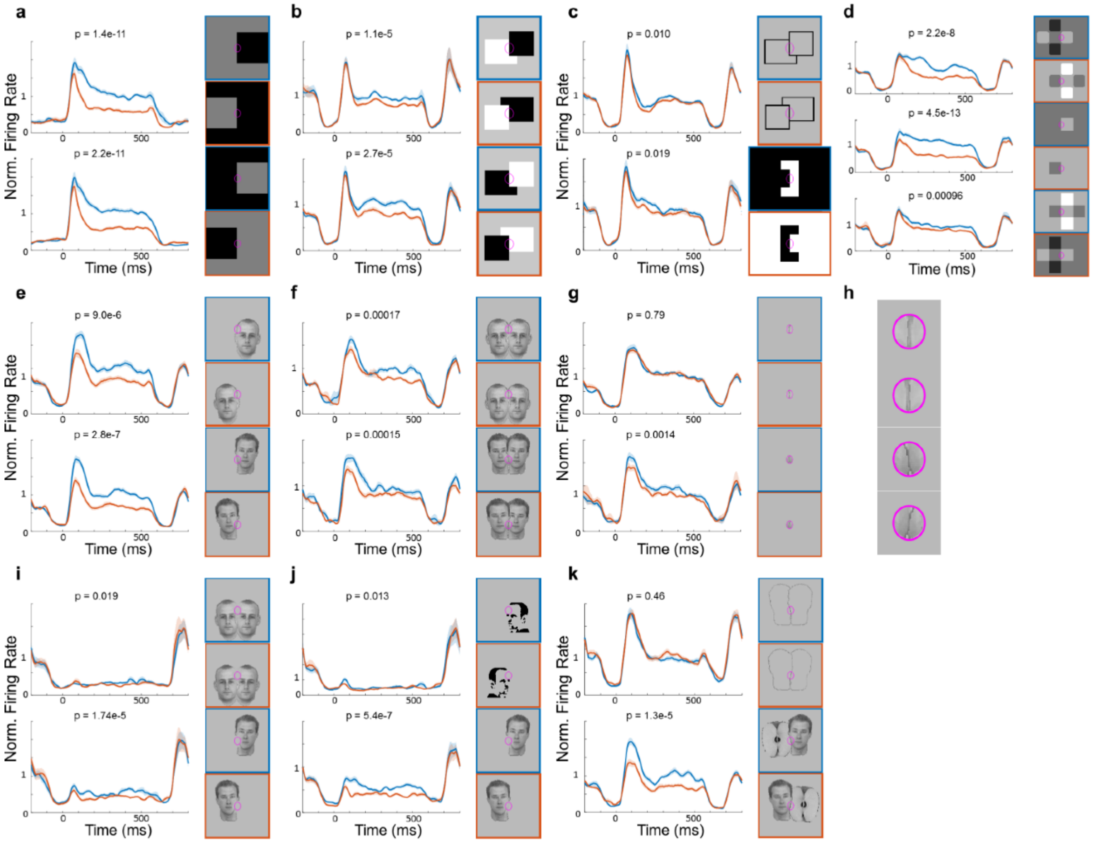


Figure III-5: Responses of border-ownership cells to simple and natural stimuli. *The PSTHs on the left of each plot show the population mean responses (±1 standard error mean, shaded region) of 50 selected border-ownership cells to stimuli shown on the right outlined with the corresponding color. Both artificial and natural stimuli were rotated and positioned such that the central edge (indicated by*

*purple ellipse) was in the preferred orientation and centered on the receptive field of each cell. Preferred side of figure of border-ownership cells was determined based on the response to simple luminance squares (shown in a) alone. Blue colors correspond to conditions where the figure was on the preferred side and red are matched conditions, where the stimulus looks identical/similar in the receptive field (purple ellipse) but the figure is on the non-preferred side. The 50 cells were determined to be consistent border-ownership cells based on their responses to simple, artificial stimuli alone. P-values were determined using two-tailed paired t-tests. (**a**) Population mean responses to luminance squares. A higher response to the preferred side is expected as the preferred side was determined based on the response to the luminance square. (**b**) Population mean responses to occlusion stimuli. (**c**) Population mean responses to occluding outlines and C-shapes. (**d**) Population mean responses to four squares (top), single squares (middle), and transparent overlay (bottom). (**e**) Population mean responses to single faces. (**f**) Population mean responses to overlapping faces. (**g**) Responses to isolated edge stimulus of faces alone. This stimulus was generated by removing the part outside the receptive field for stimuli in (f). (**h**) Magnified version of stimuli in (g). (**i**) Population mean responses to occluding and single faces with illusory contour, where the local stimulus inside the receptive field was removed. (**j**) Population mean responses Mooney faces (top) and to stimulus on bottom of (f) but with low-contrast border, where the contrast was reduced locally inside the receptive field (bottom). (**k**) Population mean responses to outlines of occluding faces and faces occluding an apple.*

Next, we asked how this population of border-ownership cells that responded consistently to the side-of-figure of artificial stimuli responded to a variety of natural stimuli. We found that the side-of-figure selectivity to artificial stimuli generalized to natural stimuli (Fig. III-5e-k): when presenting the edge of a single face in the receptive field, the average response of border-ownership cells was significantly higher when the face was on the preferred side compared to when it was presented mirror-symmetrically on the non-preferred side. Interestingly, the consistency of responses was even stronger and more significant than for most artificial stimuli based on which the cells were selected for. Moreover, the response was higher when the foreground face of two overlapping faces was presented on the preferred side. We also presented the isolated local part of the overlapping face stimulus to determine how much border-ownership information was given in the local edge alone. For the first pair of overlapping faces (Fig. III-5f top) we had chosen a central edge that was located on the forehead and fairly straight (see Fig. III-5g for the isolated stimulus and Fig. III-5h for an enlarged version). For this pair, we could find no significant response difference for the isolated local stimulus alone. For the second pair, we chose a more convex edge at the chin and the local stimulus exhibited a T-junction at the bottom of the stimulus (Fig. III-5f-h bottom). This local stimulus by contrast did evoke a significant response difference that was consistent with the border-ownership of the cells. We next presented faces with illusory contours, namely Mooney faces and the natural faces mentioned above with the local edge over the receptive field deleted. For each of these stimuli, we ensured that within the receptive field, the stimulus was identical to the background presented during the 150 ms OFF time. As can be seen from Fig. III-5i-j, the response amplitude was much weaker when deleting the local stimulus, but the response difference was still significant and consistent with the border-ownership selectivity. When presenting a stimulus with the local edge not completely deleted but strongly reduced in contrast (Fig. III-5j bottom), the amplitude was in between the amplitudes to the full face and the locally deleted face, and again the response difference was consistent with the border-ownership. We tested whether the response difference could be evoked by presenting the mere outlines of the overlapping faces (Fig. III-5k, top), and found the outlines alone did not evoke a significant response difference. The border-ownership selectivity for natural stimuli was not limited to faces, as a face occluding an apple also evoked consistent border-ownership (Fig. III-5k, bottom). Note, however, that for this stimulus, we cannot exclude that the response difference was due to the contrast difference between the two different objects. Fig. III-4b shows the proportions of cells that were consistent and significantly tuned to a given number of natural stimuli. Interestingly, among the 50 cells that were selected based on consistency for artificial stimuli (green line), there was a higher proportion of cells that was consistent and significant across many natural stimuli. This shows that the consistency across artificial stimuli generalizes to natural stimuli. Note also that for the natural stimuli in Table 1, the percentage of inconsistent cells substantially decreases when taking into account only the 50 cells with highest modulation indices averaged across artificial stimuli.

We next further examined the relationship between side-of-figure preference for artificial stimuli and natural stimuli (Fig. III-6). In general, the modulation indices for artificial and natural stimuli showed small but strongly significant correlations, indicating that cells signaling the side-of-figure of artificial stimuli correctly also tend to correctly signal the side-of-figure of natural stimuli. Yet, there is considerable residual variance of natural stimuli modulation indices left, indicating that not every cell is consistent for every stimulus, but the correct side-of-figure needs to be inferred from the population activity. Both the luminance square modulation index and average modulation index across artificial stimuli were positively correlated with natural stimulus modulation indices; however, for both single and overlapping full faces, the average modulation index across artificial stimuli explained almost twice as much variance as the modulation index for luminance squares alone. Indeed, we found that for most natural stimuli, the average modulation index across artificial stimuli explained more variance than the modulation indices of each individual artificial stimulus. This suggests that for these natural stimuli, using a battery of artificial stimuli is a better predictor for whether a cell is a border-ownership cell and will respond consistently to these natural stimuli than just using the luminance square standard test alone. In contrast, for the stimuli with ambiguous contours, the average modulation index across artificial stimuli was a worse predictor than the modulation index for the standard test alone. This could suggest that the populations of border-ownership cells that represent the side-of-figure of illusory contours might not always carry complete information about the correct segmentation of certain artificial stimulus conditions. Note, however, that, for the single ambiguous face (Fig. III-5i, bottom) the modulation index averaged across artificial stimuli actually explained more variance than the single luminance square modulation index. It is noteworthy that the single ambiguous face was the stimulus for which none of the top 50 cells was inconsistent (Table 1), whereas for artificial stimuli there tended to be many inconsistent cells. For other natural stimuli, the number of inconsistent cells was also low compared to artificial stimuli. There were three out of 50 cells that were significantly tuned to the single full face stimulus but inconsistent with the luminance square stimulus, and no cells that were significantly tuned to the single full face stimulus and inconsistent with the single ambiguous face.

Justifying our choice to focus on the most consistent cells, the consistency between border-ownership selectivity for artificial and natural stimuli was greatly weakened when we analyzed all cells regardless of their consistency across artificial stimuli (compare marginal distributions using all cells (gray bars) with the distributions using only the top 50 cells (blue bars) in Fig. III-6). When we included all cells, the population average could not significantly determine the border-ownership of occluding faces. For single faces, there was still a consistent border-ownership signal left even when averaging across all cells but it was less significant than when averaging across the top cells (despite larger sample size).

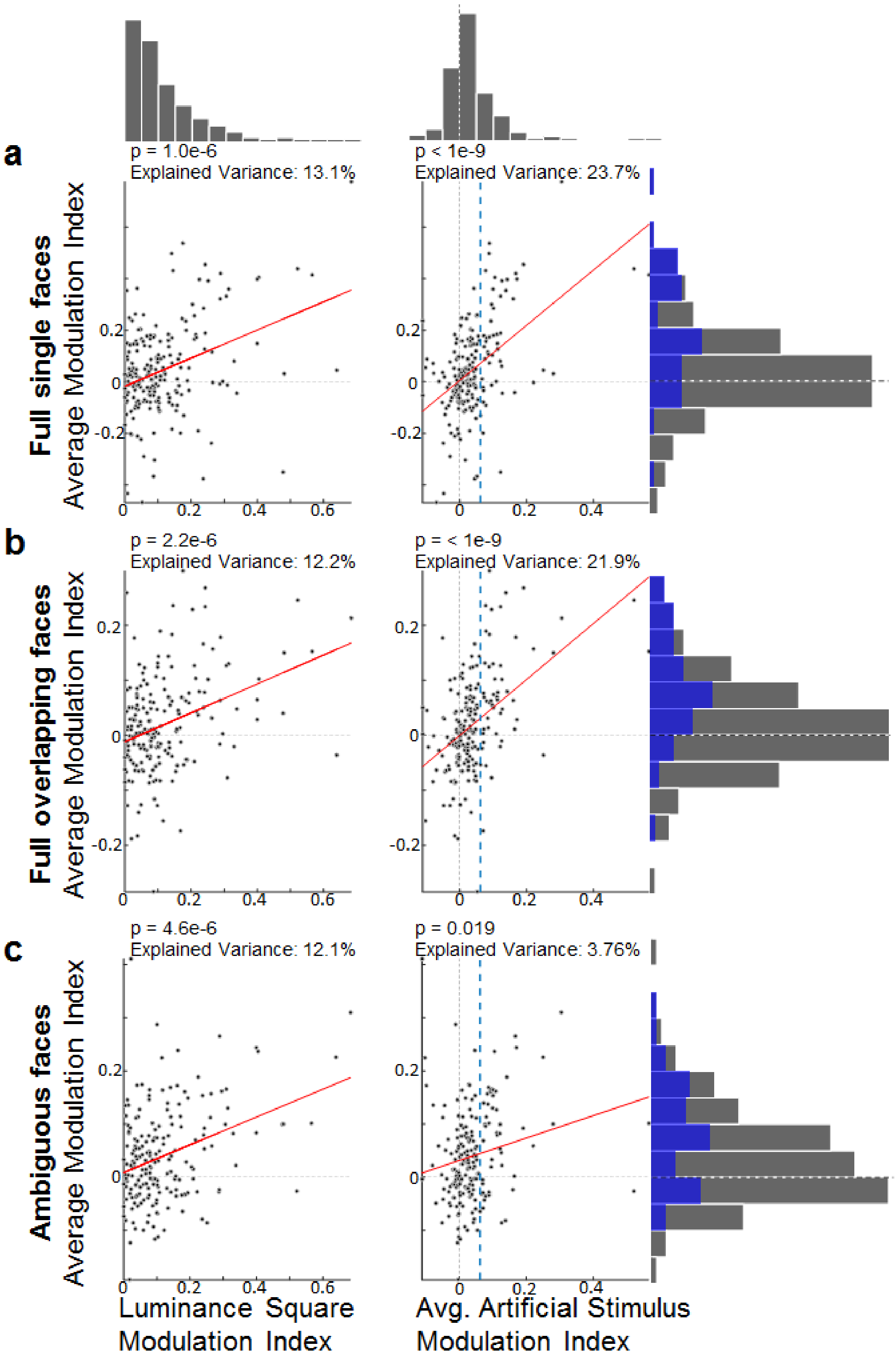

a
p = 1.0e-6
Explained Variance: 13.1%
p < 1e-9
Explained Variance: 23.7%
Full single faces
Average Modulation Index
0.2
0
-0.2
0
0.2
0.4
0.6
b
p = 2.2e-6
Explained Variance: 12.2%
p = < 1e-9
Explained Variance: 21.9%
Full overlapping faces
Average Modulation Index
0.2
0
-0.2
c
p = 4.6e-6
Explained Variance: 12.1%
p = 0.019
Explained Variance: 3.76%
Ambiguous faces
Average Modulation Index
0.2
0
Luminance Square Modulation Index
Avg. Artificial Stimulus Modulation Index

Figure III-6: Correlations between modulation indices of artificial stimuli and modulation indices of natural stimuli. *The six scatter plots show how responses to the side-of-figure of artificial stimuli correlate with responses to the side-of-figure of natural stimuli. The scatter plots on the left have the modulation indices of the single luminance square on the x-axis while the scatter plots on the right have the average modulation index across all artificial stimuli on the x-axis. Both types of modulation indices are plotted against the average modulation index across a class of natural stimuli on the y-axis: The first category (**a**) includes full single faces shown in Fig. III-5e, the second category (**b**) contains overlapping faces (Fig. III-5f) and the third category (**c**) includes faces with ambiguous contours (Fig. III-5i,j). Red lines indicate Standard Model I regression lines, i.e., the ordinary least squares fit of the abscissa to the ordinate. P-values were computed by comparing a transform of the Pearson correlation coefficient to a Student's t-distribution. Using a permutation test on the Pearson correlation coefficient yielded qualitatively the same results. The vertical dashed blue line in the right plots indicates the average modulation index which was the threshold for being chosen for the 50 selected cells used in Fig. III-5. The histograms at the top and on the right show the marginal distributions of average modulation indices for luminance squares, artificial stimuli and natural stimuli, respectively, across all 201 cells (blue histograms show modulation indices for the 50 selected cells).*

We compared latencies of the border-ownership signal for the full face stimulus and the ambiguous face stimulus. Fig. III-7a shows raster plots of an example cell's responses to full and ambiguous face stimuli. As can be seen from the difference of the PSTHs for the preferred and non-preferred condition, the border-ownership signal for the stimulus with illusory contour is smaller and delayed. As discussed in the previous paragraph, the 50 cells most consistent for artificial stimuli formed a population that signaled border-ownership of the ambiguous stimulus reliably. Thus, we compared latencies for the single full face stimulus and corresponding single ambiguous stimulus for the 17 out of 50 cells that were consistently and significantly border-ownership selective for both stimuli. The latency for the full stimulus was on average 65 ms and significantly shorter than the latency to the stimulus with illusory contour, which was on average 100 ms.

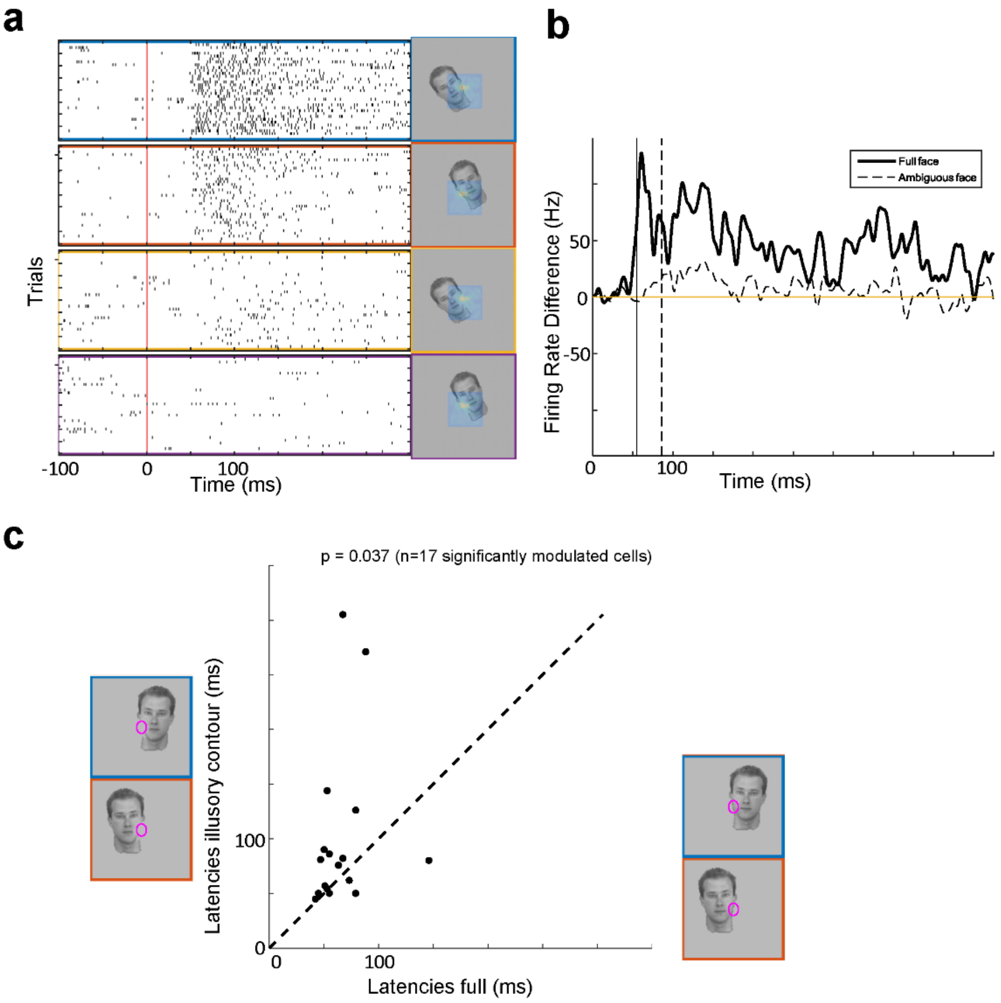


Figure III-7: Latency differences between full stimuli and stimuli with illusory contours. *(**a**) Raster plots of example cell responses to full faces presented on its preferred side (blue) and non-preferred side (red), respectively, and responses to the stimulus with locally deleted edge (yellow and purple). The preferred side was defined by the cell's responses to luminance squares. (**b**) Difference of the PSTHs for preferred side and non-preferred side, for the full face (solid line) and the face*

*with locally deleted edge (dashed line). The vertical solid and dashed lines show the latencies for full and ambiguous faces, respectively, defined as the time the difference reaches half of its peak value. (**c**) Population scatter plot of latencies of border-ownership signal of full face stimulus shown at the x-axis against border-ownership latency for the stimulus with locally deleted edge shown at the y-axis. P-value was computed with a paired Wilcoxon rank sum test.*

Discussion. We were interested in how border-ownership cells, which have been shown to respond consistently to the side-of-figure of artificial stimuli, respond to natural stimuli and stimuli with ambiguous contours. In our study, we probed responses of a large number of V2 and V3 cells for border-ownership selectivity across a large battery of different stimuli, both artificial and natural, enabling us to rigorously assess the extent to which each cell showed consistent tuning across different stimulus conditions. As can be seen from Table III-1, we found many cells that were ostensibly border-ownership cells based on selectivity to the side-of-figure of single luminance squares but responded inconsistently to other artificial stimuli. Indeed, there was not a single border-ownership cell (out of 201 tested) that was significantly tuned and consistent to every single artificial stimulus. The considerable number of partially inconsistent cells in Table III-1 contradicts the simplistic concept of the perfect border-ownership cell, despite the intuitive appeal of a single cell explaining a variety of perceptual phenomena. Instead, it seems that single border-ownership cells carry incomplete information about figure-ground segmentation for only a subset of conditions. Thus, in order for the brain to reliably determine the correct segmentation of the scene, it needs to average the activity of multiple border-ownership cells that each carry information about border-ownership in different situations. Indeed, we found that by averaging across the most consistent cells, it was possible to get a population signal which was reliable across artificial stimuli.

We found that this population signal also consistently signaled the side-of-figure of a battery of natural face and object stimuli. This further supports the notion that border-ownership cells play a vital role for segmenting the visual scene into objects. For rather simple stimuli, such as the single faces, one could argue that the consistent response might simply be caused by asymmetric receptive fields such that the cells prefer more complex texture on one side of their receptive field center. In contrast, the consistent response to overlapping faces, which are visually very similar in both conditions, indicates that border-ownership cells are indeed inferring the side of the foreground object. For the pair of overlapping faces at the bottom of Fig. III-5g, where the local edge is convex and contains a telling T-junction, even the isolated local stimulus evoked a consistent, though smaller, response difference (Fig. III-5h, bottom). This is consistent with psychophysical and computational evidence by Fowlkes et al. (2007) that the local bottom-up cues of borders in natural scenes are in many situations enough to decide border-ownership. On the other hand, when the local stimulus is ambiguous and does not evoke a significant response difference (Fig. III-5h, top), global context cues can help to decide border-ownership (Fig. III-5g, top). To test whether shape of the object is sufficient for border-ownership cells to determine the side-of-figure, we presented the outlines of the overlapping faces alone. This did not evoke a significant response difference, which is consistent with the subjective experience that the border-ownership of the outlines alone is ambiguous without the texture and other features of the face. These results suggest that border-ownership cells integrate a variety of the object’s features, including local cues, as well as shape and texture outside the receptive field.

Illusory contours have been a major subject of study for figure-ground segmentation in psychological literature (Heitger et al., 1994), and von der Heydt et. al (1984) found neurons in V2 that responded to illusory contours; however, the amplitude was smaller and the latency 10 ms longer compared to real contours (von der Heydt and Peterhans, 1989). The existence of such cells in V2 and also

V1 was confirmed by several other studies (von der Heydt and Peterhans, 1989; Grosof et al., 1993; Sheth et al., 1996; Lee and Nguyen, 2001; Ramsden et al., 2001), and Bakin et al. (2000) found that they also respond to depth-defined illusory contours. Yet, up to now, it has not been known whether border-ownership cells consistently signal the side-of-figure for illusory contours (Kogo and Wagemans, 2013), although there have been several computational models (Finkel and Sajda, 1992; Sajda and Finkel, 1992; Kogo et al., 2010) where border-ownership and illusory contours both emerge from a dynamic network computing figure-ground organization. Recently, there has also been a discussion paper (Kogo and Wagemans, 2013) suggesting the intertwinedness of illusory contours with border-ownership, which received many commentaries. Among them, von der Heydt (2013) argued that illusory contours and border ownership might be represented by distinct populations in V2. Here, we have shown that border-ownership cells do consistently signal the side-of-figure for illusory contours even though local edge information is missing. The number of cells significantly signaling the correct side-of-figure for illusory contours was slightly lower than for full contours (31 out of 50 significant cells for full stimulus vs. 20 out of 50 significant cells for ambiguous stimulus, unpaired t-test, $p<0.05$), which may be partly due to the reduced amplitude of responses to illusory contours in general. Among the 50 cells most consistent for artificial stimuli, there were only 3 cells that were significantly tuned for the full face stimulus but inconsistent for the luminance square, and no cells that were significantly and inconsistently tuned to the side-of-figure of the ambiguous stimulus and the luminance square. It is surprising that the ambiguous face stimulus turned out to be the stimulus where no cell signaled inconsistent border-ownership. A possible explanation is that the local stimulus in the classical receptive field causes the transient response, which can lead to errors for the unambiguous stimuli, whereas for the ambiguous face stimulus, the border-ownership signal is evoked entirely by feedback from context (Gilbert and Li, 2013), which is more reliable. Analogously to the onset responses of illusory contours studied by von der Heydt and Peterhans (1989), the response differences signaling border-ownership of illusory contours were also smaller and delayed by about 30 ms compared to the full stimuli.

Previously, the responses of border-ownership cells have been modeled as a result of pure feedforward operations (Heitger et al., 1994; Sakai and Nishimura, 2006; Supèr et al., 2010), of intra-areal dynamics (Baek and Sajda, 2005; Zhaoping, 2005), and of feedforward and feedback interactions between areas (Craft et al., 2007; Jehee et al., 2007; Kogo et al., 2010), respectively. Our findings falsify pure feedforward models as they predict the same latencies of border-ownership signals irrespective of the stimulus. Instead, recurrent (Lamme and Roelfsema, 2000) connections might need to be utilized to resolve ambiguous scenes. It is conceivable that the latency of the border-ownership signal is increased because intra-areal network dynamics or feedforward-feedback interactions between V2 and V4 require more iterations to resolve the border-ownership. Note, however, that at least for the Mooney face, which evokes a significant and consistent border-ownership signal (Fig. III-5j), it seems unlikely that contour completion mechanisms building from low-level cues would be sufficient to infer the illusory boundary of the face, but instead knowledge about face shape appears to be required. Assuming that the border-ownership signal that emerges around 100 ms for the full face originates from feedback from V4, it is also possible that the later signal for the border-ownership of the illusory contour arises from feedback from an area later in the hierarchy, e.g., posterior IT. Such cortico-cortical feedback loop interactions have been shown to exist between V1 and V4 by (Chen et al., 2014). The most posterior face patch PL has latencies of 80 ms to distinguish faces from objects (Issa and DiCarlo, 2012) and is thus a possible candidate feedback source.

Previously, there has been a debate (Vecera and Farah, 1997; Vecera and O'reilly, 1998; Peterson, 1999) on whether segmentation precedes recognition or vice versa. Our results (Fig. III-7) suggest the possibility of a third alternative that is consistent with models of Bayesian inference (Rao and Ballard, 1999; Lee and Mumford, 2003; Yuille and Kersten, 2006). According to this hypothesis, initially, retinotopic areas try to segment the scene into regions corresponding to objects based on low-level cues, potentially using a combination of feedforward mechanisms utilizing local cues as T-junctions, intra-areal, and inter-areal dynamics. However, since segmentation of natural scenes is inherently ambiguous based on low-level cues and not every edge is an object border, this initial segmentation wave can only make guesses about which regions correspond to objects and relay these regions to IT. IT then attempts to recognize objects in the hypothesized regions and can accept or falsify the hypotheses by exciting or inhibiting border-ownership cells via feedback. In this way, IT would generate a representation of object surfaces in retinotopic cortex. Note, however, that the found latencies are merely suggestive evidence, and simultaneous recordings and perturbations of multiple areas will be necessary to dissect the exact mechanisms.

Overall, we found that many ostensible border-ownership cells, as determined by the single luminance square, turned out to be inconsistent for one or more stimuli when presented with a larger battery of artificial stimuli, and not a single cell showed consistent border-ownership preference across all stimulus conditions tested. This emphasizes the necessity for future studies to present a larger set of stimuli in order to identify the most consistent border-ownership cells and the need to use a population code for decoding segmentation. Importantly, the population of border-ownership cells that was consistent across most artificial stimuli could also reliably segment both natural face and object stimuli and, with some delay, even ambiguous stimuli where local edge information was completely missing, which suggests that border-ownership cells integrate both local, low-level cues and global, high-level object cues to segment the visual scene.   By exploiting new techniques such as population calcium imaging, optogenetics, and simultaneous recordings in retinotopic and IT cortex, future work might be able to reveal how border ownership cell populations are read out across different stimulus configurations.

Materials and Methods. All animal procedures used in this study complied with local and NIH guidelines. Two male rhesus macaques were implanted with MR-compatible head posts and trained to maintain fixation on a dot for a juice reward.

Targeting. Since the interest of this study lies explicitly in recording from border-ownership cells rather than an exhaustive analysis of V2/V3, our electrode targeting was guided by fMRI. Monkeys were scanned in a 3T TIM (Siemens) magnet. Scanning procedures were the same as described in Tsao et al. (2006), Freiwald and Tsao (2010), and Ohayon and Tsao (2012). For functional MRI, monkeys passively viewed stimuli on a screen. MION contrast agent was injected to improve signal to noise ratio. In order to identify V2 and V3, we first mapped retinotopy by presenting horizontal and vertical checkerboard wedges and defined area boundaries based on horizontal and vertical meridians. Within V2 and V3, we targeted areas with high functional activation in response to border-rich disparity-defined checkerboard stimulus vs. a full field changing disparity stimulus (Tsao et al., 2003b) in order to increase the yield of recorded border-ownership cells (Fig. III-2a), since Qiu and Von Der Heydt (2005) had previously found that a majority of border-ownership cells are also selective to the side-of-figure of stereo-defined edges, which were abundant in the former stimulus. In addition, we were guided by anatomical landmarks, and targets were confined to the banks and fundus of the lunate sulcus. For monkey J, we did not get good signal in the fMRI and therefore targeted the same anatomical locations that yielded high activation in monkey T. We found slightly less border-ownership cells than in monkey T (see Results). Placement of recording chambers and electrode trajectories towards the targeted regions were planned with the software Planner (Ohayon and Tsao, 2012). In monkey T, we recorded from the right hemisphere and in monkey J from the left hemisphere.

Fixation. Monkeys were head fixed and passively viewed a screen in a dark room. A small fixation spot (0.25° in diameter) was presented in the center of the screen and eye position was monitored using an iScan system. Monkeys were rewarded with juice for maintaining fixation every 2-3 seconds.

Electrophysiology. Tungsten electrodes (FHC) with 1 MΩ impedance were used for recording. Custom grids were printed and guide tubes were cut to extend 2 mm below the dura. Electrodes were advanced using an oil hydraulic microdrive (Narishige). Neural signals were recorded using a MAP system (Plexon). Local field potentials (LFPs) were filtered at 0.7–300 Hz, and single units and multi-units were filtered at 0.15– 8 kHz and recorded at 40 kHz.

Online Data Analysis. Spikes were isolated and sorted online using the box-method of the SortClient (Plexon). Initially, approximate receptive field location was determined by manually sweeping a small blinking square (0.2 º) across the screen. Based on this approximate location, receptive fields were mapped by computing the spike-triggered average (STA) in response to a random stimulus of size 8º that was centered on the hand-mapped location (Pack et al., 2003). The random stimulus was a series of images alternating at 100 ms that consisted of a grey background and two squares of size 0.5º appearing at random positions, with one of the squares being white and the other square randomly chosen to be either black or white. Subsequently, a 2-dimensional Gaussian was fitted to the spike-triggered average (STA) of the stimulus to determine the position and size of subsequent stimuli. Receptive field maps were also computed by considering either the only black squares or the white squares alone and yielded similar receptive fields. Next, moving sine wave gratings were presented, and the preferred orientation of the cell was determined based on the sine grating orientation that evoked the highest response. For all subsequent stimuli, the central edge (indicated by a purple ellipse in the figures) was adjusted to the position and size of the receptive field and rotated to match the preferred orientation. We recorded a total of 545 cells in monkey T and 121 cells in monkey J. We subjectively assessed for each recorded site whether it contained promising border-ownership cells based on clear receptive fields, clear orientation preferences, and consistent responses to the side-of figure of luminance squares and for a total of 298 out of 545 recorded cells proceeded to present a battery of border-ownership tests consisting of both artificial stimuli and natural stimuli of faces (note that some of these cells were included because they were recorded simultaneously as ones which passed the subjective assessment, but were not themselves subjectively assessed). In the offline analysis, further cells were excluded based on unclear STAs or insufficient samples of responses, yielding a total of 201 valid cells (see Offline Data Analysis). Artificial stimuli consisted of stimuli that had been shown by von der Heydt and colleagues to evoke consistent border-ownership signals, including the standard test with single luminance squares (Zhou et al. 2000), two occluding squares (Zhou et al. 2000), occluding outlines, C-shapes, and squares that evoke the perception of either four single squares or a transparent overlay (Qiu et al. 2007). To further verify the correct mapping of the receptive field and test position invariance within the receptive field, we also performed position tests by sliding the luminance square across 11 positions orthogonal to its preferred orientation. Natural stimuli consisted of single faces, overlapping faces, the isolated local edge of overlapping faces alone, faces with local edge deleted, Mooney faces, outlines of overlapping faces, and faces occluding apples. The whole battery of artificial and natural stimuli is shown in Fig. III-5. We chose to use mostly faces for the natural stimuli as they represent a natural, complex, high level object category that is of strong behavioral and social relevance and with which monkeys have extensive experience. Also, the existence of face-selective regions in IT opens up the possibility to examine interactions between object representations in retinotopic and IT cortex (Tsao et al., 2008a). Stimuli were presented for 500 ms ON time and 150 ms OFF time. To correct for delays of the screen, we used a photodiode that detected the onset and offset of the stimuli. The photodiode's output was fed into the recording system and later used to synchronize the onset of the stimulus and the neurophysiological data during offline analysis.

Offline Data Analysis. Spikes were re-sorted offline using OfflineSorter (Plexon). Trials in which monkeys broke fixation were discarded (using a 1° eccentricity fixation window). We discarded cells with insufficient number of trials or spikes (<500 or <1500 total spikes for standard test and natural stimulus set, respectively), and cells that either had an unclear STA were not centered on the receptive field or failed the position test (i.e., when shifting the stimulus as in Fig. III-2e, the peak response was not inside the receptive field), leaving a total of 201 cells. Peristimulus time histograms (PSTHs) were smoothed with a Gaussian kernel. For Table III-1, which shows the consistency of cells for artificial and natural stimuli shown in Fig. III-5, we determined the proportion of side-of-figure selective cells for different stimuli using the same method as in (Zhou et al., 2000): side-of-figure selectivity of a cell to a given stimulus was computed using a two-way anova on side-of-figure and contrast polarity on the average firing rate from 0 to 500 ms during trials of different conditions and using an unpaired t-test in case of only one contrast polarity. Unless stated otherwise, a cell was deemed significantly selective to the side-of-figure of a stimulus if p<0.01 and consistent if the modulation index for the given stimulus had the same sign as for the standard test of single luminance squares. Modulation indices for pairs of matched stimuli were computed as $M_{s1,s2} = \frac{R_{s1}-R_{s2}}{R_{s1}+R_{s2}}$, where $R_{stimulus}$ is the firing rate of the cell in response to the stimulus averaged over the 500 ms from onset to offset of stimulus presentation. We defined the preferred side-of-figure of each cell based on the average modulation index for the standard luminance squares shown in Fig III-2. Based on this preferred side, we computed the average modulation index across all pairs of matched artificial stimuli (blue and red conditions in Fig. III-5a-d). This average modulation index is positive if the border-ownership coding across artificial stimuli is consistent with the border-ownership selectivity for the simple luminance square and negative if it is inconsistent. We selected the 50 most consistent cells with the highest average modulation indices across artificial stimuli to compute the population average responses shown in Fig. III-5. The results did not depend qualitatively on the exact number of selected cells. Before averaging, each PSTH was normalized by the average response from 0 ms to 500 ms after stimulus onset across the shown stimuli. The p-values shown on top of each PSTH were computed using two-sided paired t-tests on the normalized average responses from 0 to 500 ms after stimulus onset across trials of the 50 neurons for the preferred vs. non-preferred side. For the scatter plots in Fig. III-6, we computed the average modulation index across the two local contrast conditions of the luminance square and the average modulation index across all artificial stimuli, and computed the correlation with modulation indices for natural stimuli. Finally, we compared latencies of the border-ownership signal for full face stimuli and ambiguous face stimuli with illusory contours. Traditionally, latencies have been computed as the time when the signal first significantly exceeds baseline fluctuations (Maunsell and Gibson, 1992; Kiani et al., 2005). However, since the response amplitudes were much lower for stimuli with illusory contours, we were worried that this definition might be biased toward longer latencies for stimuli with lower signal-to-noise ratios. Thus, we (1) only included the subset of border-ownership cells in the analysis that showed significantly stronger responses to the preferred side for both the full stimulus and the stimulus with illusory contour ($p<0.05$, Welch test). The preferred side was determined by responses to the luminance squares. (2) We used a half-peak measurement as latency (Zhou et al., 2000): we computed the difference of smoothed (SD: 9 ms) PSTHs for preferred and non-preferred side and defined latency as the first time that the difference reached half of the maximum difference across the 500 ms of the stimulus ON time. (3) We repeated the analysis with a change point measure (Sugihara et al., 2011), which fits a piecewise linear function consisting of two lines to the cumulative difference PSTH and defines the latency as the point where the first leg, which is fixed as 0, transitions to the second leg, which qualitatively confirmed our results. All analysis was performed using Matlab (Mathworks).

Segmentation patches

Summarizing the findings of the previous section, individual cells recorded from random locations in V2 and V3 were incapable of reliably signaling border-ownership across different types of stimuli. This was bad news of our original goal of recording from clear signals encoding segmentation and recognition and to study their interaction during switches of conscious percept. Therefore, we asked (1) whether there are specialized modules where neurons do reliably encode the figure-ground segmentation of a scene. Perhaps, we could apply an approach similar to the one that led to the discovery of face cells (see Chapter II) and employ fMRI paradigms to identify functional clusters of cells that encode a larger amount of figure-ground information. And (2) we asked whether we can use a decoding approach and use responses of a population of neurons recorded while the monkey viewed different types of figure-ground stimuli to decode the figure-ground segmentation of each image. Methods for fMRI and electrophysiology were the same as described in the section above.

Identification of fMRI-defined segmentation hotspots. We hypothesized that if modules involved in figure-ground segmentation exist, then stimuli containing figures on a background would trigger these processes and evoke activation in those modules, whereas stimuli without figures would not (since there is nothing to segment). We thus presented stimuli containing figures on a background as well corresponding control stimuli containing only a background while monkeys were being functionally scanned (Fig III-8). We used either texture, luminance, motion, or disparity to define each of the figures and their respective background stimuli. Analyzing the difference in responses, we found hot spots of activation in V2, V3, V4, and V4A that were activated more strongly by stimuli containing figures than pure background stimuli. Interestingly, these activations by figures were very similar for each of the modalities, i.e., the overlap of each of the contrasts – whether defined by texture, luminance, motion, or disparity – was remarkably strong. This overlap in hot spot activation was not be explained simply by higher signal-to-noise ratio in those regions: as a control, we performed an fMRI experiment comparing colored stimuli to greyscale stimuli. This contrast showed very little overlap with the segmentation-related hot spots, except for one patch in V4v, raising the possibility that segmentation-related modules may be largely separated from color-processing machinery.

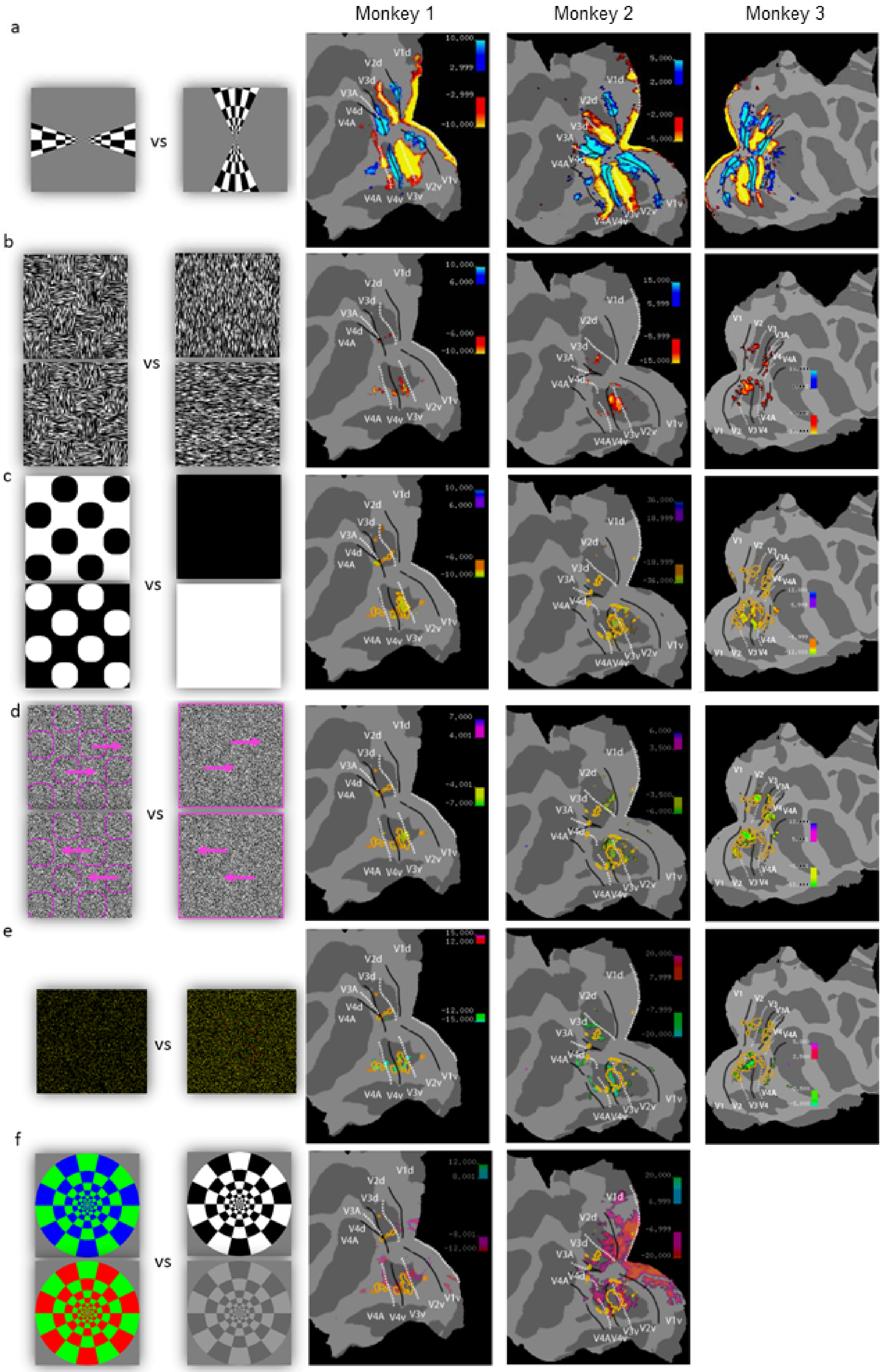

Monkey 1
Monkey 2
Monkey 3
a
vs
b
vs
c
vs
d
vs
e
vs
f
vs
V1d
V2d
V3d
V3A
V4d
V4A
V1v
V2v
V3v
V4v
V4A
V1
V2
V3
V3A
V4
10,000
2.999
-2.999
-10,000
5,000
2,000
-2,000
-5,000

***Figure III-8: Segmentation hotspots overlap across modalities.** We first mapped retinotopy to determine boundaries of visual areas V1-V4 (**a**) in three monkeys (three columns). We then presented texture-defined stimuli and compared activations to control stimuli with full-field background texture which revealed regions of activations located mostly in V3 and V4 (**b**). Outlines of these patches are shown in orange in the flat maps below. Stimuli containing figures defined by luminance (**c**), motion (**d**), or disparity (**e**) caused activations in overlapping regions, while activations evoked by color stimuli had little overlap (**f**).*

Cluster of consistent cells in segmentation hot spots. We targeted these segmentation hotspots with electrophysiological recordings to compare figure-ground information encoded by neurons inside these regions with neurons from outside control regions, including those in the patches defined by color, which had previously been called globs and were suggested to be responsible for color processing (Conway et al., 2007). During electrophysiological recordings, we showed stimuli similar to the fMRI stimuli, in that they were defined by either texture, luminance, motion, or disparity. For each modality, either the center or edge of a single square was presented on the receptive field of a cell, and as a control, we also showed background stimuli without a square. While in general, consistent segmentation cells were rare, we did find a cluster of cells inside a segmentation hot spot where we, reproducibly across days, found cells that were consistent across all four modalities. The location of this cluster also overlapped with disparity columns that Adams and Zeki (2001) had studied inside the lunate fundus. Similarly to their findings, we found that as we advanced our electrode along the lunate fundus, cells changed from being tuned to near disparity to being tuned to far disparity (Fig. III-9a). An example cell from the cluster of consistent cells is shown in Fig. III-9b. Inside this cluster, cells that were consistent across all four modalities were much easier to find than anywhere outside the cluster (Fig. III-10).

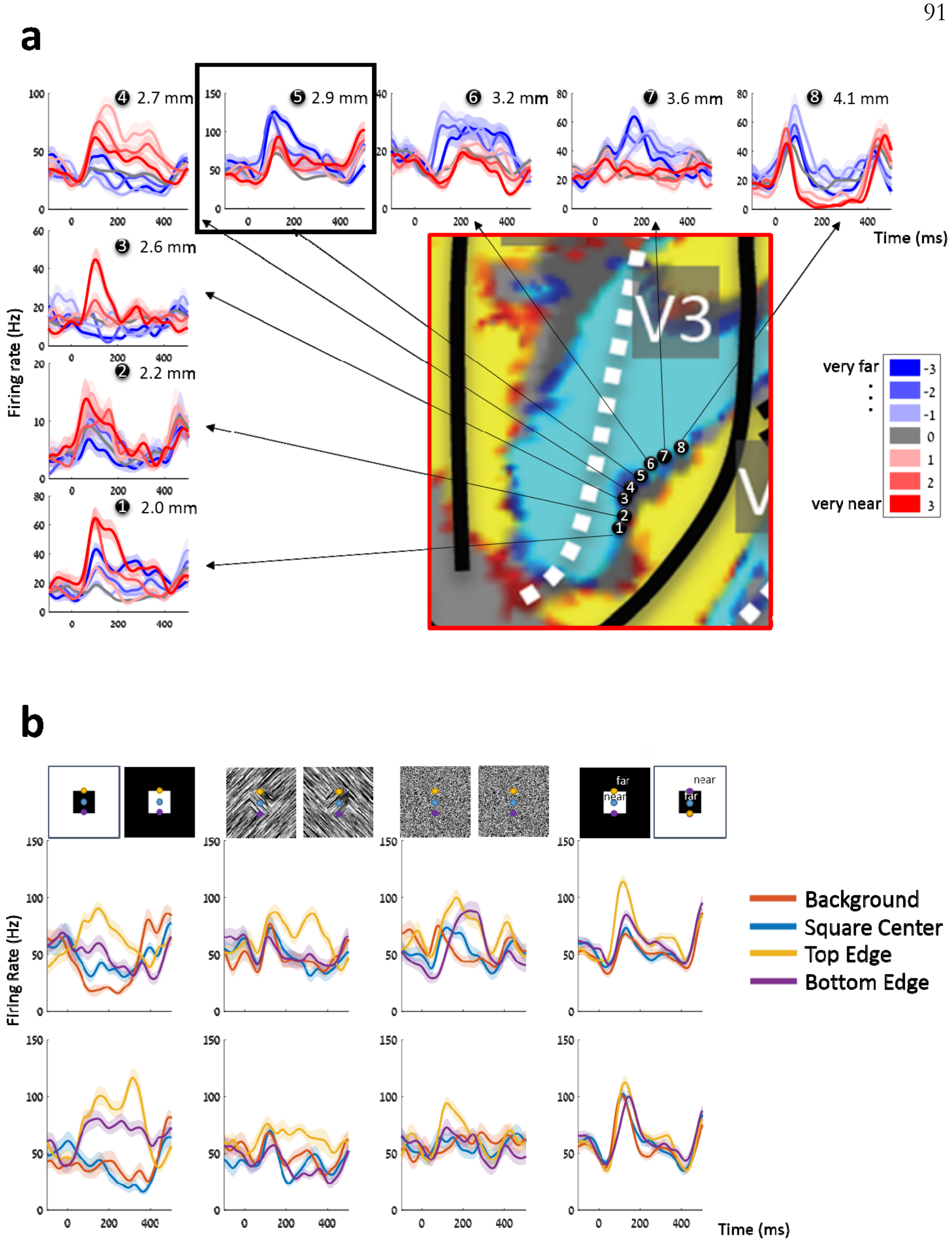


Figure III-9: A cluster of consistent cells inside of segmentation hot spot / disparity columns. *We implanted a chamber almost horizontally so we could*

*make long penetrations in the fundus of the lunate sulcus along V3. Shown in the center of (**a**) is a part of a flatmap showing V3 with retinotopy superimposed and 8 recording locations that we recorded along the lunate fundus. Shown around the flatmaps are PSTHs representing the disparity tuning (from far, blue; to near, red). As the electrode was advanced, neurons changed tuning from preferring near disparity to preferring far disparity. The cell recorded at site 5, which also happened to be inside a segmentation hot spot (not shown), was consistently tuned to the side-of-figure of an edge as shown in the PSTHs in (**b**). This neuron fired more strongly when the top edge of a square was presented over its receptive field, regardless of whether the square was defined by luminance, texture, motion, or disparity.*

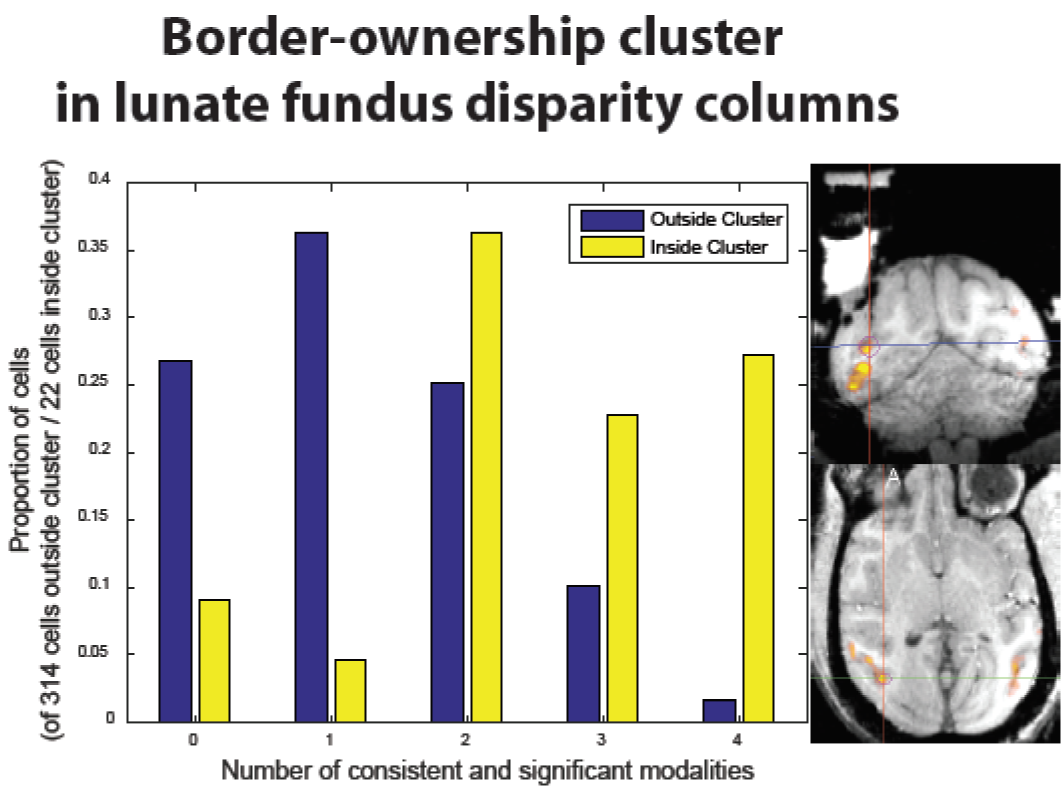


***Figure III-10: Inside the cluster, cells are remarkably consistent**. Left: Histograms comparing the proportion of cells that were significant and consistent*

across 1, 2, 3, or 4 modalities inside and outside of the cluster. Right: Location of the cluster, with disparity-defined segmentation hot spot fMRI contrast overlaid.

Decoding accuracies are higher in segmentation hot spots. We next characterized the amount of figure-ground information tuned in population of cells from specific structures of the visual cortex, such as segmentation-activated hot spots, globs (color modules), and control regions outside of segmentation hot spots and globs. To this end, we trained a linear classifier on responses of neurons and tried to predict whether a given region of an image was part of a figure, background, or a top edge/bottom edge of a figure (chance level thus being 25% decoding accuracy). Inside segmentation-activated regions as measured by fMRI we found higher percentages of cells tuned to borders compared to outside these regions for all modalities and also higher decoding accuracies reaching around 60% when using 50 neurons for decoding (Fig. III-11). Decoding accuracies in control regions were notably lower, and decoding accuracies in globs were particularly low.

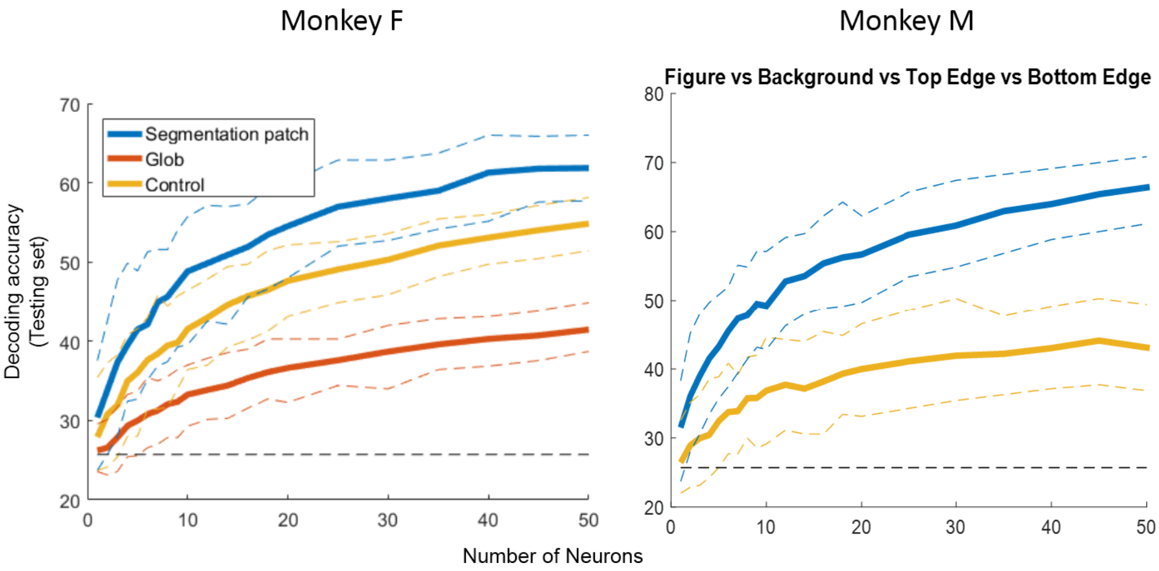


Figure III-11: Decoding figure-ground information. *Decoding accuracy and 95% confidence intervals (y-axis) against numbers of neurons used (x-axis) for two*

monkeys. To determine how much information about figure-ground segmentation is encoded in different patches of visual cortex, we used a linear classifier to extract figure-ground information from responses of either neurons from segmentation-activated regions (blue), glob neurons (red), and neurons from control regions outside of both segmentation-activated regions and globs. Squares defined by different modalities (luminance, texture, motion, disparity) were presented at different positions relative to the receptive field (square center, top edge, bottom edge, or full background stimulus without the square). A linear decoder was trained on a single-trial basis to classify across stimuli whether a figure, a top edge, a bottom edge, or background was over the receptive field. For computing confidence intervals, we repeated decoding 100 times, and in each of the 100 iterations, a random subset of neurons and trials were selected.

Chapter IV

MOONEY FACES

We advise the reader to have a brief glance at Fig. IV-1 and see if he or she recognizes the objects in the image. Moreover, he or she should try to infer the figure-ground assignment, i.e., determine which regions in the image are foreground and which are background. Most subjects will have a hard time doing this. Next, turn the page upside-down and try to recognize the object in Fig. IV-1 again. Now, it should be quite easy to see that the image depicts two faces, namely of a woman (my advisor) and of a monkey. Also, it becomes easy to infer where foreground and where background is. Importantly, even after turning the page back to its normal position (and the faces to their inverted position), the ability to recognize and segment the two stimuli persists. Hence, Mooney faces are switchable (although the switch can arguably only happen once), and may be usable to study whether face patches signal switches in percept. Given the previous chapter, another reason that makes the Mooney face an attractive potential stimulus for us was that Mooney faces cause the perception of subjective contours where the contours of a real, non-degraded face would usually be. Experiments described above suggested that border-ownership cells also encode to which side these illusory contours belong.

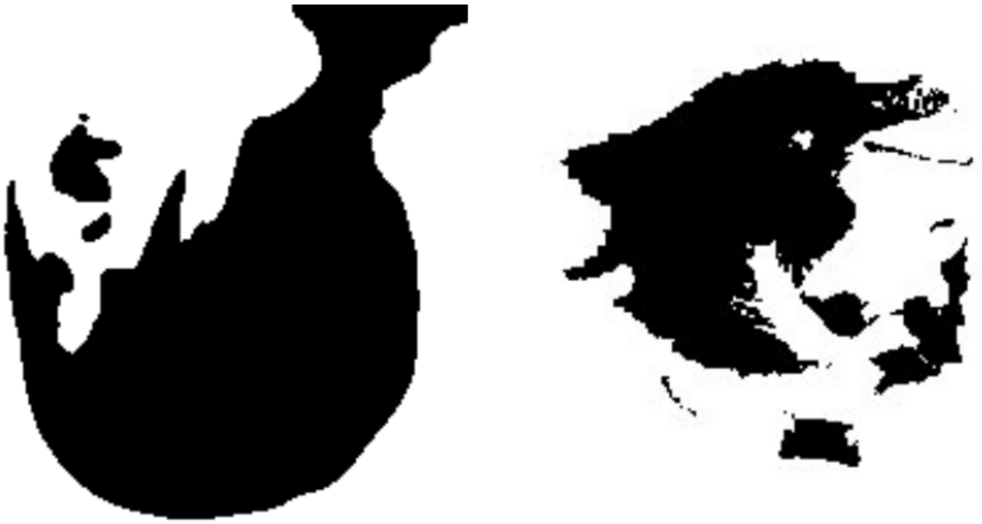


Figure IV-1: Example Mooney face stimuli*. If the reader finds it hard to recognize these stimuli after trying for a while, note that the Mooney stimuli have been inverted. Putting the page upside-down makes it easy to recognize them for most human subjects.*

While humans are very good at recognizing upright Mooney faces from the age of 18 months (Doi et al., 2009), in most monkeys, Mooney face stimuli activated face patches only very weakly (Moeller et al., 2017). Strikingly, one monkey in our colony, which anecdotally was also deemed one of the smarter monkeys, showed face patch activation to Mooney faces just as strong as to real faces (see Figure 6 in Moeller et al. (2017)). This raised the possibility that most monkeys may have difficulty recognizing Mooney faces as depicting faces in the same way that humans have difficulty recognizing upside-down Mooney faces. The outlier suggested it may be possible to make monkeys *see* the Mooney faces in the same way that turning the page upside-down helped the reader. Thus, we decided to find out why monkeys showed low activation to Mooney faces and if we could make them respond strongly to them. Two ideas we pursued for making face patches respond to Mooney faces were (1) to figure out what the difference is between Mooney faces and cartoon faces (which cause strong face responses) and (2) to animate the Mooney faces. We investigated the two questions with electrophysiology and fMRI, respectively, using the same methods as described in Chapter III.

For the first question, we recorded from 85 face-selective cells in face patch MF (Fig. IV-2). While face cells in face patch MF responded to Mooney faces slightly more than to real objects or Mooney objects, the response was very small, confirming results cited above. Indeed, the baseline-subtracted peak response to real faces was about four times as large as the response to Mooney faces (Fig. IV-2a). This could not be simply explained by the fact that Mooney faces were not fully realistic faces: even for simplistic cartoon faces, the base-line subtracted peak response was almost three times as large as for the Mooney face. Thus, we asked what made this big difference between cartoon faces and Mooney faces. For this purpose, we created chimeras by exchanging face parts of cartoon and Mooney faces (Fig. IV-2b). We found that the outline of the cartoon face had by far the strongest influence on responses – cells responded to every second stimulus, which corresponds exactly to the stimuli that had the outline of the cartoon face present. We confirmed the importance of having an outline through chimeras between real faces and Mooney faces (data not shown).

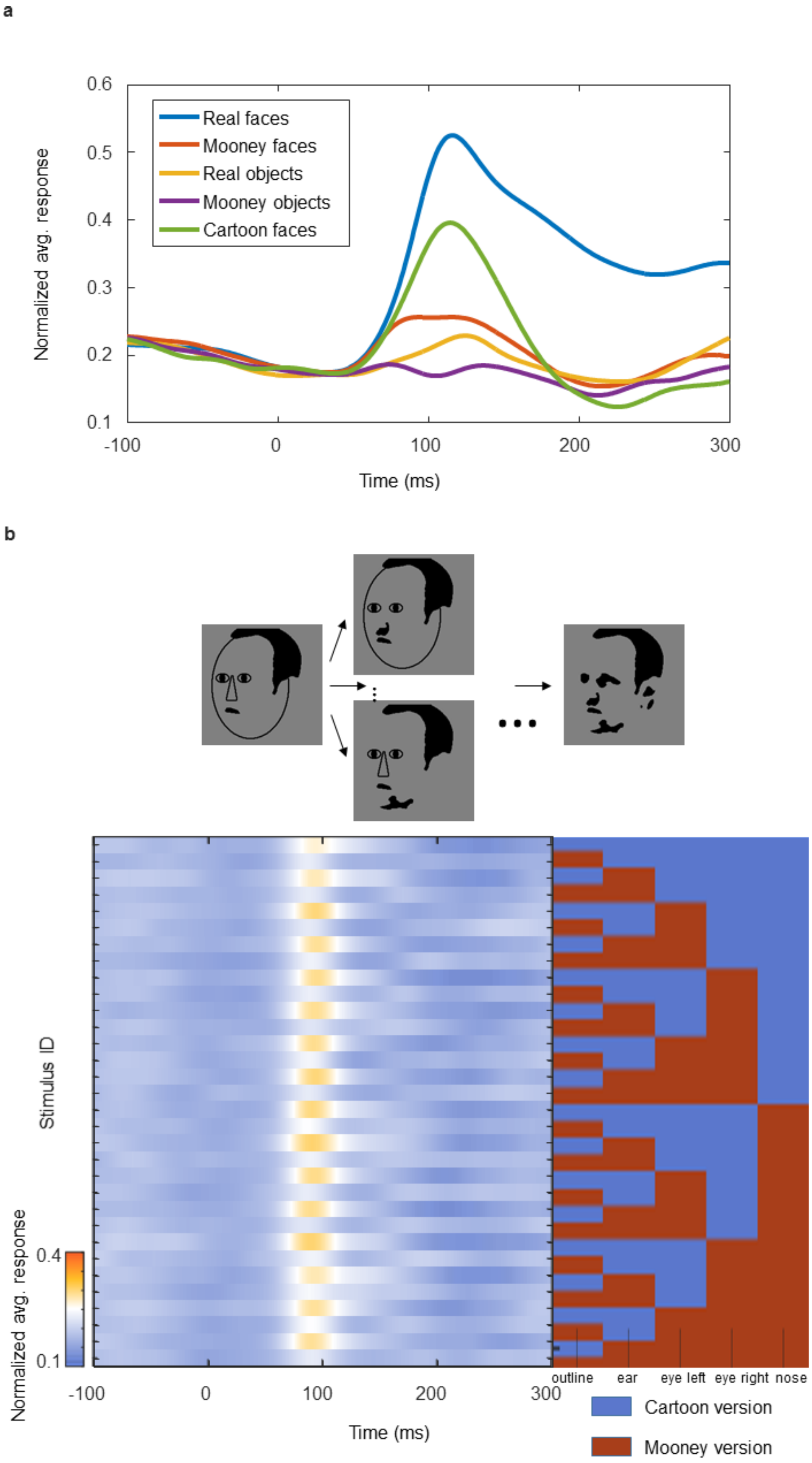

a
Normalized avg. response
0.6
0.5
0.4
0.3
0.2
0.1
Real faces
Mooney faces
Real objects
Mooney objects
Cartoon faces
-100
0
100
200
300
Time (ms)
b
Stimulus ID
Normalized avg. response
0.4
0.1
-100
0
100
200
300
Time (ms)
outline
ear
eye left
eye right
nose
Cartoon version
Mooney version

Figure IV-2: Responses to Mooney faces can be increased by adding an outline. *(**a**) Responses averaged across 85 MF cells (normalized) to real faces, Mooney versions of the same faces, real objects, Mooney versions of the same objects, and cartoon faces. (**b**) Top: Schematic illustrating the process of generating cartoon – Mooney chimeras by replacing one part (outline, ear, eye left, eye right, nose) at a time. Bottom: Response time courses averaged across 85 MF cells to each of the $2^5 = 32$ possible combinations of using either cartoon version or Mooney version for each part. Diagram on the right indicates for each row which parts were Mooney and which parts were cartoon for the respective stimulus.*

For the second experiment, we made a shot-in-the-dark hypothesis that motion cues of a naturally moving Mooney face may help to increase face patch responses to Mooney faces. Thus, we created movie stimuli of real faces and objects moving around in a natural way, and also generated Mooney versions of the same stimuli. Mooney versions were created by smoothing the image of each frame and thresholding image intensity to create an image made of only two tones. We showed these movies, as well as corresponding static images, while monkeys were in the scanner to see how animation influences fMRI responses in face patches. Consistent with single-unit results described above we found that responses to static Mooney faces were significantly weaker than to real faces (Fig.IV-3a). Interestingly, animating the Mooney face had a remarkable effect, eliciting responses similar in magnitude to real faces. Importantly, this could not be explained simply by the motion energy of the stimuli: animating Mooney object stimuli did not increase their response. In a second experiment, in the same monkey, we additionally showed upside-down static or animated Mooney faces (Fig. IV-3b). The animation-induced response increase for upside-down Mooney faces was stronger than for Mooney objects but weaker than for upright Mooney faces. We confirmed the main findings across face patches in a total of three monkeys (data not shown).

Concluding this chapter, we were able to make face patches respond to Mooney faces by adding an outline or animating them. One interpretation of these results frames them in terms of figure-ground segmentation: perhaps, because Mooney faces lack many of their contours (at least physically – they may still be perceived subjectively when the Mooney face is recognized), segmentation processes fail, and hence the features of a face cannot be grouped together as belonging to the same surface. This may impair object recognition processes. Adding a physical outline clearly helps the segmentation process, and natural motion can also serve as powerful cues to aid segmentation. Further experiments, such as using disparity to define the surface of the Mooney face, may gather further support for this hypothesis.

Importantly, unlike the perceptual hysteresis in humans, face patches did not appear to “learn” from this experience. For example, for the second day of the fMRI experiment, after the monkey had seen both static Mooney faces and their animated counterparts for hours, responses to static Mooney faces remained low (Fig. IV-3b). Hence, we were not able to exploit Mooney faces as a switchable stimulus for our interaction experiments. Different possible hypotheses are consistent with these findings: for example, whether the monkey learned that a Mooney face represents a face or not, or already knew it from the beginning, it is possible that this knowledge is purely semantic and encoded in higher cognitive areas, not the face patches. Alternatively, the monkey may have switched from not perceiving a face to perceiving a face, but face patches may not represent the monkey’s percept but be mere passive filters of the visual input. This underscored for us that, before studying the interaction between brain areas, we would first need to figure out whether face patches actually represent conscious percept or not. We decided to definitely answer this question using binocular rivalry as a paradigm, as described in the next half of this thesis.

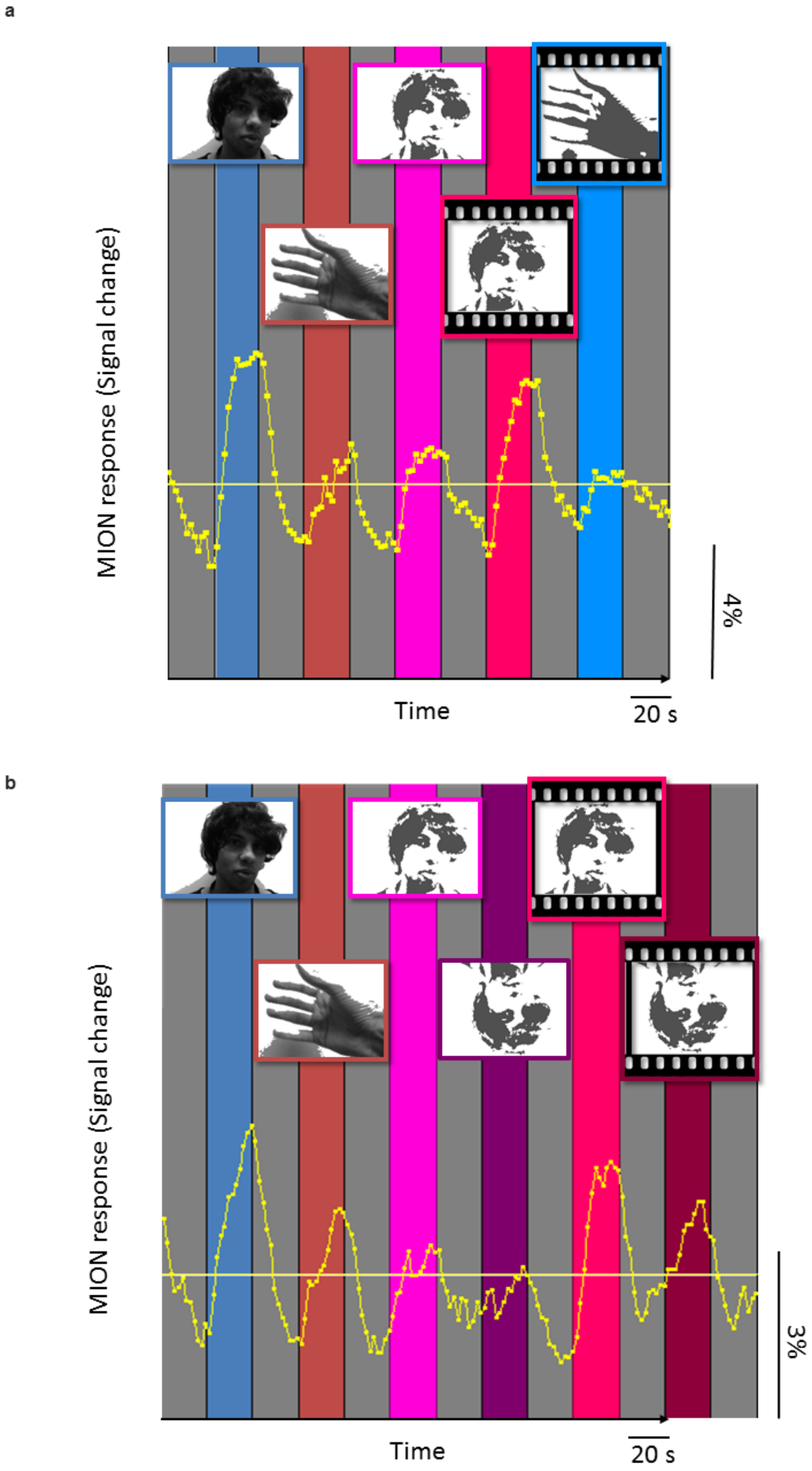

a
MION response (Signal change)
Time
20 s
4%
b
MION response (Signal change)
Time
20 s
3%

Figure IV-3: Responses to Mooney faces can be increased through animation. *(**a**) Response time course, averaged across 14 runs, of the fMRI signal. Values above base line mean activation, values below baseline mean suppression. Each dot of the line plot corresponds to one functional volume taken (repetition time TR = 2 seconds). During each block, consisting of 12 TRs, either static real faces, static real objects, static Mooney faces, Mooney face movies, or Mooney object movies were presented, interleaved with blank periods (grey). Each category consisted of 8 stimuli. (**b**) Same as (a), but using different blocks on a different day. Images presented during blocks were real faces, real objects, static upright Mooney faces, static upside-down Mooney faces, upright Mooney face movies, and upside-down Mooney face movies.*

Chapter V

A PHILOSOPHICAL TREATISE ON CONSCIOUSNESS

Definitions of consciousness

Consciousness is the biggest remaining mystery of nature. While processes previously viewed as magical such as life have been elucidated by biology, real scientific progress on consciousness is scarce. To date, it remains almost completely unclear how consciousness, which can be defined as subjective experience or what it feels like to, e.g., see the color the red, maps to the physical world. While Aristotle asserted that consciousness must of course reside in the heart (Hicks, 2015), these days it is believed that rather the brain has something to do with consciousness. Thus, if looking to understand consciousness, one commonly glances hopingly at neuroscience and psychology.

One has to distinguish between generic consciousness and specific contents of consciousness. Generic consciousness is the capability of having subjective experience at all, i.e., the difference between an awake individual and a deeply sleeping or anesthetized person. In neuroscience, the neural correlates of generic consciousness are studied by measuring brain states awake vs. asleep or anesthetized, or perturbing specific brain regions and see whether this renders the patient unconscious. However, we currently have no way of specifically turning off consciousness without also turning off a myriad of other processes, such as language, planned movement, higher-order thoughts, decision making, etc. Hence, it is currently not possible distill the neural correlates of generic consciousness from these other processes, which is the reason why I will focus on specific consciousness instead. Specific consciousness is about the contents of conscious experience, i.e., the subjective perception of the color red vs. the color green. Here, we have a greater chance to disentangle the subjective experience of seeing red vs.

not from all the other processes that come with it, and can thus hold more hope to find the true neural correlates of experience.

Another distinction one needs to make is between phenomenal and access consciousness (Block, 1995). Phenomenal consciousness addresses the subjective, first-person view of conscious experience, or what it feels like to experience something, which is a private, non-physical sensation that cannot be conveyed directly to other entities. On the other hand, access consciousness addresses the third-person view of consciousness that can be observed behaviorally, i.e., whether a person has access to the perceived information in the sense that he can utilize or report it. To make it clear from the beginning, I do not believe that neuroscience has a very good shot at the hard problem of fully solving phenomenal consciousness, as I do not think that neuroscience will eliminate the conceivability of philosophical zombies. Even if we have a complete understanding and detailed description of the neural underpinnings of consciousness, it is still coherently conceivable for the same physical system to exist but exist in the dark, without anyone experiencing what it is like to be in that system. But this conceivability shall not demoralize us, as we can still get very close to a full understanding of consciousness: in the best-case scenario, we will be able to find the states or processes of the brain that empirically coincide exactly with specific contents of consciousness and only with consciousness, i.e., the provided description of brain processes is not modulated by other, confounding factors. If this is the case, we will be able to exactly predict, constrained only by the limits of language, what your current phenomenal experience is, and by perturbing exactly these processes, you will be able to reproducibly experience any exact conscious percept desired. The process description will be sufficient in the sense that if we clamp the process and perturb another part of the brain, your conscious percept will be unaffected and minimal in the sense that if we miss perturbing a specific aspect of the process, we will not be able to provoke the full experience. So even though zombies will still be conceivable, the empirically discovered relationship between consciousness and the physical world will be more like a fundamental law, such as gravity in physics. After all, we can also conceive of a world with the same particles having mass, etc. as in the real world, but gravity happens to be absent. This conceivability does not imply that the discovered physical laws of gravitational force have not contributed to a deep understanding of gravity.

Access consciousness seems much easier to solve than phenomenal consciousness. We can easily conceive finding the pathway by which, e.g., light is picked up by the photoreceptors, processed through several stages, and leading eventually to the motor actions of reporting the perceived information. Indeed, almost all studies trying to find neural correlates of conscious percepts, discussed below, depend on the subjects' report of what they are perceiving. A critical distinction to be made here though is that the subject consciously perceiving something is not identical to the subject reporting his conscious percept. Indeed, one could decide to never report any experience and would still be conscious (cf. dreaming state and locked-in patients). Being conscious of something means generally (but not always) that one is *capable* of reporting it but does not have to. In almost all studies in neuroscience, the conscious percept exactly coincides with the report of the percept during the experiment, so it is impossible to tell whether the neural correlates correlate with consciousness or the report itself. To make things worse, there are several other confounding factors, such as attention, decision-making, and differences in the physical stimulus that need to be controlled besides report. Thus, when trying to extract consciousness from the neurophysiological soup, one needs to be extra careful not to pick up the things associated with other processes that tend to stick to it.

In the following sections, I will summarize results from the neuroscience literature on three major paradigms, flash suppression, backward masking, and binocular rivalry, that are being used to distill the neural correlates of a conscious percept. I will discuss to which degree the results agree and compare how effective the three paradigms are in dissociating consciousness from other confounding factors. I will constrain the scope on results from electrophysiology. A lot of related work has been done using psychophysics and functional magnetic resonance imaging (fMRI). However, psychophysics tells us comparatively little about the brain, and the results from fMRI happen to disagree dramatically with electrophysiology regarding where the conscious percept is represented, with fMRI generally overestimating the extent to which it is represented in the brain (e.g., Tong and Engel (2001)). Since electrophysiology is considered the gold standard as to what neurons represent, I choose to focus on electrophysiology here, although the discrepancies between fMRI and single-unit electrophysiology are an interesting subject for discussion. In the end, I will argue for the main thesis of this paper that no convincing dissociation between conscious percept and confounding factors has been performed in hitherto experiments and will propose an experiment utilizing a no-report paradigm to distill the representation of specific contents of consciousness.

Review of previous paradigms for specific consciousness

How does one find the neural correlates of a conscious percept? A moderate correlation is not enough: if one flashes light at a subject, the subject will have the subjective experience of brightness and coincidently the retina will be activated with each flash. Despite the retina's firing being correlated with the subjective percept, one cannot claim it to be the neural correlate of consciousness, because it is not an exact correlate. That is, there is a lot of sensory input that we are not conscious of but that is represented in the retina. Thus, we need to dissociate what activations correlate exactly with the conscious percept and what activations are just driven by the mere physical input. To dissociate the physical input from the conscious percept, we want to find paradigms where the physical input is fixed but the conscious percept changes over time or is different over repeated presentations. The studies discussed below aim to achieve this using three different strategies:

In backward masking, a stimulus is presented so briefly that it can only perceived on some trials but not on others. One can compare neural activity when the stimulus is reported to be seen vs. not seen with the physical input being identical in both conditions. Flash suppression shows a target stimulus to one eye, while flashing a salient distractor stimulus in the other eye. This way, the target stimulus is suppressed by the distractor and is only perceived once it recovers from suppression after a variable amount of time. Binocular rivalry is an example of bistable stimuli, such as the famous face-vase illusion (Rubin, 1980), where the percept spontaneously switches between two percepts. As for flash suppression, two different stimuli are presented to the two eyes, but the conscious percept keeps stochastically alternating between the images in the two eyes.

Backward Masking. In the ‘‘backward masking’’ paradigm (Breitmeyer et al., 1984) an image of e.g., a face or a tool is briefly presented, shortly followed by a visual ‘‘mask,’’ a meaningless picture aimed at disrupting the recognition process. This can render the target image to be at the threshold of recognition, i.e., even for the same stimulus, the subject can only sometimes recognize the image.

In areas associated with object recognition, firing rates (KovAcs et al., 1995) and gamma power (Fisch et al., 2009) were increased if the subject detected an object in a backward masking task. In the higher-level medial temporal lobe, which transforms percepts into memories, firing rates and gamma power of the local field potential (LFP) were also increased if the subject detected the stimulus, but there was an even earlier deflection in delta power before the increase in firing rates, and the differences in LFP power were more predictive than the changes in firing rate (Quiroga et al., 2008; Rey et al., 2014).

Supèr et al. (2001) did not use a mask but showed a stimulus containing a texture-defined figure to monkeys that they had to saccade to. Comparing trials where the monkey succeeded saccading to the figure with trials where he failed, the early response in primary visual cortex (V1) was identical, but the late response (>100 ms) was suppressed on miss trials, suggesting an influence of feedback to V1 on conscious perception.

Flash suppression. During flash suppression, a target stimulus is presented to one eye, while a second distractor stimulus is flashed to the other eye, thereby suppressing the target stimulus. In single flash suppression (Wolfe, 1984), the target stimulus is first presented to one eye, and subsequently both stimuli are shown binocularly. The flashing of the distractor stimulus thereby usually suppresses the target stimulus, which has not changed. A second paradigm called continuous flash suppression (Tsuchiya and Koch, 2005) keeps flashing the distractor stimulus at around 10 Hz, which yields prolonged suppression of the target stimulus up to minutes. Unlike binocular rivalry discussed below, where the stimuli are usually balanced, in flash suppression the distractor stimulus is usually rich in structure, such as Mondrian patterns, to increase strength of suppression. In a third paradigm, called generalized flash suppression (Wilke et al., 2003), a monocular target is suppressed by a binocular distractor in the *surround* of the target.

In V1, generalized flash suppression caused no change in firing rate but modulated oscillations of the LFP in the 9-30 Hz and gamma range. In the lateral geniculate nucleus (LGN), the thalamic structure providing input to V1, no modulations were observed, but firing rates in pulvinar, a higher-level thalamic structure associated with attention, did change according to stimulus visibility during generalized flash suppression (Wilke et al., 2009). Neural activity in prefrontal cortex has also been shown to correlate the with conscious percept during single flash suppression (Panagiotaropoulos et al., 2012; Kapoor et al., 2018).

Bistable stimuli and binocular rivalry. If one stimulus is presented to the left eye, and a completely different stimulus is presented to the other eye, rather than seeing a superimposition of the two, one will usually perceive only one of them, followed by the other, alternating stochastically in an endless dance. Since the physical input is fixed throughout, the fluctuations in conscious percept are entirely internally generated, providing a good entry point for isolating the neural correlates of specific contents of consciousness (Tong et al., 2006; Blake et al., 2014). Binocular rivalry is a special case of bistable stimuli, such as the face-vase illusion or ambiguous motion, and has the advantage of increased freedom in choosing the two images that the percept oscillates between.

Logothetis and Schall (1989) trained monkeys to report the perceived motion direction while viewing bistable motion patterns, and found that MT was modulated by the reported motion direction. During binocular rivalry, it was found that only as little as 20% of neurons in V1 represent the reported conscious percept and that the percentage increases as one goes up the visual cortex hierarchy to V2 and V4 (Leopold and Logothetis, 1996), disagreeing with the fMRI literature that did find strong correlations with the conscious percept even in V1 (Polonsky et al., 2000a; Tong and Engel, 2001). In cat V1, Fries et al. (1997) claimed that binocular rivalry increased synchrony and regularity of gamma oscillations but not average firing rate. In the lower-level, thalamic lateral geniculate nucleus, which provides input to V1, no modulation by binocular rivalry was observed (Lehky and Maunsell, 1996). In the higher-level object recognition area inferotemporal (IT) cortex, 90% of cells in encoded reported percept (Sheinberg and Logothetis, 1997). Activity in the human medial temporal lobe and frontal cortex, which are higher-level than IT, has also been shown to reflect the reported conscious percept (Gelbard-Sagiv et al., 2018).

Confounding factors

When trying to distill the neural correlates of consciousness, one needs to carefully control for several confounding factors that tend to co-occur with changes of the conscious percept, such as changes in the physical stimulus, attention, report, and memory. Otherwise, one may be measuring the neural correlates of these confounding factors rather than of the conscious percept. In the following sections, we will discuss why each of these confounding factors is a problem and to which degree the three paradigms above avoid them.

Physical stimulus differences. As discussed above, even the retina will respond to changes in the physical input, but we cannot infer from this that the retina is a neural correlate of consciousness. Hence, one should try to keep the stimulus fixed and only change the conscious percept of it. In binocular rivalry and backward masking, the stimulus is constant, or at least the same across repeated trials, respectively. For flash suppression, there are two ways to collect and analyze the data. (1) One can keep repeating the same flash suppression experiment and ask subjects to report when the suppressed stimulus is released from suppression and enters the conscious percept. Since the time of suppression is stochastic, one can compare activity when the suppressed stimulus starts being perceived at different times. (2) In so-called no-report paradigms, the subject is just passively viewing the stimulus, and one compares two stimulus conditions, one where the target is suppressed by a flashing distractor, and the other where the target is shown without the distractor. Here, the problem arises that the physical input is very different in the two cases. Consider for example the results above, that the response to a preferred stimulus was decreased if a distractor was flashed in the other eye, compared to when the preferred stimulus was presented alone. We know from studies in IT, that if a distractor is presented, then the response to the preferred stimulus is decreased even if the preferred stimulus is perfectly visible, e.g., when the distractor is shown next to or transparently overlaid on the preferred stimulus (Bao and Tsao, 2018). Thus, it is expected for the distractor to decrease the response, not because the conscious percept of the preferred stimulus is suppressed, but just because of the physical stimulus difference. Alternatively, in the single flash paradigm, people sometimes compare responses to showing only the preferred stimulus first followed by binocular presentation, to responses to showing only the non-preferred stimulus first followed by binocular presentation. Upon the binocular presentation, they claim that the response is stronger if the preferred stimulus is shown second, in accordance with perception. However, this is also expected because of the transient responses of neurons: the firing rate of a neuron rapidly adapts to a stationary stimulus, so if the preferred stimulus is shown first, the firing rate has naturally already declined by the time the second stimulus is flashed.

Thus, current no-report paradigms of flash suppression have severe problems with differences in physical input, whereas backward masking and binocular rivalry come out unharmed.

Attention. Attention has been suggested to be a separate, in some case even opposing, process, to consciousness (Koch and Tsuchiya, 2007). Yet, the two often go hand in hand, i.e., one becomes conscious of something when paying attention to it, so it is challenging to dissociate activity modulations caused by either of them in experimental setups. Indeed, of the confounding factors listed here, attention is probably the one most difficult to disentangle from consciousness. The paradigm having the most severe issues with attention is arguably backward masking. Whether a briefly presented stimulus is perceived or not depends a whole lot on the general level of arousal, willingness to succeed, and selective attention to the stimulus location. Thus, if we find a neural signature that correlates with whether a stimulus was perceived or not, it may just be representing the general arousal level of the subject.

For flash suppression, bottom-up attention is an important issue. For the no-report flash suppression paradigms, where one compares activity when a distractor is flashed or not, it is clear that the flashing of the salient distractor will evoke bottom-up attention, leading to the improved processing of the distractor and diminished processing of the preferred stimulus. Thus, observed modulations may represent the object-based attention to the distractor itself, the impoverished representation of the preferred stimulus due to impoverished processing, or suppression of the preferred stimulus to enhance the distractor rather than the conscious percept.

In binocular rivalry, the role of attention has been controversial. Although earlier literature claimed that binocular rivalry is not under attentional control, it is quite clear when experiencing a binocular rivalry stimulus that one has some kind of voluntary control over which stimulus one perceives (see Ooi and He (1999)). Later, it was found that compared to other bistable stimuli such as the Necker cube (Meng and Tong, 2004), the attentional control over which a stimulus is perceived is less for binocular rivalry, and one cannot hold one of the two percepts in awareness indefinitely. However, attention could be used for both types of stimuli to change the non-selective speed of switching. Later it was suggested that if attention is diverted from the rivalry stimulus by asking the subject to do a challenging different task, binocular rivalry modulations measured by EEG frequency tagging disappeared (Zhang et al., 2011). It has also been claimed that object-based attention causes dominance in binocular rivalry (Mitchell et al., 2004).

Report. Almost all of the studies mentioned above required active report of the subject, e.g., by button press, to determine what the conscious percept at a given time was. The exception is no-report flash suppression, which caused problems with physical stimulus differences, though. Thus, it is possible that all the observed modulations of neural activity represented just the act of reporting itself rather than the conscious percept. Reporting one’s conscious percept entails several processes, such as introspection about what one is perceiving, making a decision, and the motor action of e.g., pressing a button. All these processes are known to be able to cause modulations in the brain, and are thus important confounding factors. Moreover, pondering and reporting the perception of a face may bring with it feedback processes for imagining a face, which is known to activate the visual cortex (Khuvis et al., 2018). Indeed, Frässle et al. (2014) found that fMRI modulations in many brain regions that were observed when the conscious percept changed in binocular rivalry with active report vanished when the subjects did not report their percept. This is alarming and indicates that the above studies may indeed have confounded the conscious percept with the report thereof.

Memory. The three paradigms as stated so far do not have severe issues with memory, as the percept is read out immediately, and does not require prolonged storage in working or long-term memory. One could consider removing modulations caused by the report of the percept by not asking the subject to remember the switches immediately but report them later. For example, the subject could be asked to count the number of switches and report them only later, either verbally or by button press. Note that this would only control the motor action of report, and not the introspection and decision-making that would still have to occur during the binocular stimulation. However, even if the motor action can be removed from the equation, another confound would be added instead: neuronal modulations may just be representative of how well the percept was transferred into memory and thus reportable later on.

No-report paradigms

To summarize the previous section, all three paradigms turn out to be highly problematic when trying to isolate consciousness from other perception-associated processes: each of the paradigms has problems with either attention, report, or physical stimulus differences. To the rescue come novel no-report paradigms proposed by Naber et al. (2011): unlike the no-report paradigm mentioned earlier in the context of flash suppression, the physical stimulus is fixed. However, to dissociate report from conscious percept, the conscious percept is not indicated by active report, but instead inferred from other behavioral markers that happen to coincide with the percept. The first marker that can be used is the so-called optokinetic nystagmus. If a moving grating is presented to a subject, its eyes will reflexively follow the movement of the grating. It turns out that if two gratings moving in opposite directions are presented in binocular rivalry, the eye movement will follow the grating that is reported to be consciously perceived. The second marker they proposed is pupil dilation. When presented with a bright stimulus, one’s pupil will contract. When presented with a dark stimulus, one’s pupil will dilate. When presented with a bright stimulus in one eye, and a dark, incompatible stimulus such as orthogonal gratings, in the other eye, pupil dilation reflexively follows one's percept switching during rivalry. Thus, it is not necessary for the subject to actively report its percept anymore: we can infer the current percept of the subject based on these reflexes that happen to be correlated with and predictive of the conscious percept. As mentioned above, Frässle et al. (2014) used optokinetic nystagmus and pupil dilation to infer switches of conscious percept in two conditions: either with the subjects actively reporting their percept or without them doing any task. They found substantial modulation across the brain during active report, but most of the activations, including prefrontal cortex, vanished without active report. Thus, it remains to be seen which modulations of neural activity actually reflect the conscious percept as opposed to report.

Optokinetic nystagmus and pupil dilation are not ideal markers either, as they bring other potential confounds with them (Overgaard and Fazekas, 2016). Monkeys anesthetized with ketamine still show switches in optokinetic nystagmus during binocular rivalry (Leopold et al., 2002), even though they should not be conscious of the stimulus. This casts doubt on whether it is an exact correlate of the conscious percept. Furthermore, the smooth eye movements during nystagmus bring problems in physical input differences with them. Assume, stimulus 1 is moving to the left and stimulus 2 is moving to the right, and you are currently perceiving stimulus 1 and hence experience a nystagmus to the left. This means, that most of the time, stimulus 1 will be stationary on your retinas while stimulus 2 will be moving. If objects are used instead of gratings, another important confound is that the perceived object will always be on the center of your fovea, whereas the suppressed object will be moving into the periphery. This is important for foveally biased receptive fields and contralateral bias in many higher-level regions. On the other hand, pupil dilation is also modulated by arousal, uncertainty, reward prediction, and motor preparation, so it brings another stack of confounding factors with it.

Thus, I propose a third no-report paradigm for inferring the current percept: unlike the previous paradigms that are based on reflexes, in this paradigm the subject is asked to actively fixate on a fixation spot, that can appear e.g., either on the left side or the right side. In the left eye, stimulus 1, e.g., a face, is presented and in the right eye, stimulus 2, e.g., a body, is presented, leading to rivalry between the two objects. Now, the critical point is that the fixation spot in the left eye is actually different from the fixation spot in the right eye, i.e., when the left eye fixation spot is on the left side, the right eye fixation spot is on the right side, and vice versa. Perceptually, because of the rivalry, the subject will only perceive one object and one fixation spot at a given time, and the two will be linked, i.e., when perceiving the face, it will perceive the fixation spot of the left eye. Due to this link, we can predict whether he perceives face or body from where he fixates. This link between fixation position and perceived object only occurs on a single-trial basis, but as a whole is dissociated: half of the trials when perceiving a face, the fixation spot will be on the left, half of the trials it will be on the right, and the same is true for perceiving a body. Thus, any modulations we observe with changes in the inferred percept cannot be caused by the fixation spot location. In the subject's experience, fixation spot location and perceived percept are independent. It just needs to follow the one fixation spot it sees and does not need to introspect, report, or make decisions about the perceiving face or body; indeed it can just ignore the object for this task. This task thus also alleviates the attention confound, because the task directs top-down attention to the fixation point, so less object-based attention to the face or body is possible. In my subjective experience, viewing this stimulus and tracking the fixation spots does not abolish rivalry, though. I tested the paradigm in humans that simultaneously reported their percept of face or body through button press, and was able to predict the reported percept with 86%-98% accuracy. In sum, the proposed third no-report paradigm dissociates consciousness from the physical stimulus, report, memory, and, to an improved degree, from attention.

Conclusions

We have discussed to which degree neuroscience has helped or could help us to better understand specific contents of consciousness. I argued that in the best-case scenario, neuroscience has the capability of making real and meaningful progress on phenomenal consciousness. Finding the neural pathways for access consciousness is comparatively more tractable, but holds several caveats. When viewing consciousness from a third-person view, one can only indirectly infer what the subject is perceiving. Thus, one needs to be careful not to conflate the state of a stimulus being in one's consciousness, i.e., accessible by processes such as report, with the report of said stimulus itself. One needs to carefully control for differences in the physical input, attention, introspection, and report, and I argued that none of the previous electrophysiological studies has successfully done so. I thus proposed a no-report paradigm, complementary to the two no-report paradigms already proposed by Naber et al. (2011) with the goal to dissociate the mentioned confounding factors from specific consciousness. Compared to previous report and no-report paradigms, I have argued that this paradigm is better at distilling the conscious percept from the physical stimulus, report, and attention. As a caveat, I do not claim that the paradigm fully controls for attention, although the advantage is that it diverts top-down attention away from the changing percept. This paradigm can thus be used to find out whether previously reported modulations and brain regions actually reflect the conscious percept. Finding the neural correlates of specific conscious percepts will be an important step towards solving the problem of phenomenal consciousness.

Chapter VI

BINOCULAR RIVALRY

Representation of conscious percept in macaque face patches

From: Hesse and Tsao (2020), "Representation of conscious percept without report in the macaque face patch network". In : *bioRxiv*. Submitted to *eLife*.

Abstract. A powerful paradigm to identify the neural correlates of consciousness is binocular rivalry, wherein a constant visual stimulus evokes a varying conscious percept. It has recently been suggested that activity modulations observed during rivalry could represent the act of report rather than the conscious percept itself. Here, we performed single-unit recordings from face patches in macaque inferotemporal (IT) cortex using a no-report paradigm in which the animal's conscious percept was inferred from eye movements. We found high proportions of IT neurons represented the conscious percept even without active report. Population activity in single trials, measured using a new 128-site Neuropixels-like electrode, was more weakly modulated by rivalry than by physical stimulus transitions, but nevertheless allowed decoding of the changing conscious percept. These findings suggest that macaque face patches encode both the physical stimulus and the animal's conscious visual percept, and the latter encoding does not require active report.

Introduction. Having conscious experience is arguably the most important reason why it matters to us whether we are alive or dead. The question what signals in the brain reflect this conscious experience and what signals reflect obligatory processing of input regardless of conscious experience is therefore one of the most important puzzles in neuroscience. For example, activations in the retina may correlate with the conscious percept of flashing light but are arguably entirely driven by physical input, much of which never evolves into a conscious percept. Another driver of neural activity that can be confounded with signals related to conscious perception is report. Recently, it has been suggested that brain regions may correlate with conscious perception simply because they are driven by the active report of it (Aru et al., 2012; Frässle et al., 2014; Safavi et al., 2014; Tsuchiya et al., 2015; Koch et al., 2016; Overgaard and Fazekas, 2016; Tsuchiya et al., 2016; Boly et al., 2017; Block, 2019, 2020; Kapoor et al., 2020).

A paradigm known as binocular rivalry is useful for distinguishing responses related to conscious perception from those driven by obligatory processing of physical input (Tong et al., 2006; Blake et al., 2014). When two incompatible stimuli such as a face and an object are shown to the left and right eyes, respectively, one does not perceive a constant superimposition of the two, but instead one's percept alternates between face and object even though the physical input is fixed (Fig. VI-1a). Since these alternations are internally generated, they cannot be attributed to pure feedforward processing of external input.

In previous studies, researchers trained monkeys to report their percept during binocular rivalry by releasing a lever and found that the proportion of cells modulated by the reported percept increases along the visual hierarchy, with as little as 20% of cells showing modulations in V1 (Leopold and Logothetis, 1996) compared to 90% of cells showing modulations in IT (Sheinberg and Logothetis, 1997). Using fMRI, Tong et al. found that the human fusiform face area responds to reported perceptual switches (Tong et al., 1998). The reported percept also modulates activity of single units in the human medial temporal lobe and frontal cortex (Gelbard-Sagiv et al., 2018).

Although binocular rivalry isolates the conscious percept from physical input, an important confounding factor remains. In all studies cited above, the monkey or human subject always actively reported their percept by a motor response. Thus it is possible that the observed neuronal activations were due to the act of report itself, including introspection, decision making, and motor action accompanying report, rather than a switch in conscious percept. This concern was emphasized in an fMRI experiment by Frässle et al., who compared modulations in the brain with and without active report (Frässle et al., 2014). Many of the modulations observed in higher-level brain regions such as the frontal lobe disappeared when subjects did not actively report perceptual switches.

To infer the subject's percept in the absence of report, Frässle et al. used two no-report paradigms that depended on pupil size and optokinetic nystagmus, respectively. If the stimuli in the two eyes have different brightness, the subject's pupil size will vary according to the dominant percept's brightness and can thus be used to infer the percept. As a second method, Frässle et al. exploited optokinetic nystagmus. They presented gratings moving in opposite directions in the two eyes, causing the subject's eye position to reflexively follow the direction of the dominant grating.

These no-report paradigms allow accurate prediction of the subject's percept but are not free of confounds themselves (Overgaard and Fazekas, 2016). First, pupil size is known to correlate with arousal, surprise, attention, and other confounding factors (Hoeks and Levelt, 1993; Bradley et al., 2008; Preuschoff et al., 2011). Second, when optokinetic nystagmus is applied to moving non-grating stimuli such as natural objects that drive IT cortex, there will be confounding physical stimulus differences. For example, the dominant stimulus that is smoothly pursued by the subject's eyes will tend to be stationary on the subject's fovea and optimally modulate IT areas with foveal biases, while the non-dominant stimulus will be more eccentric and have increased motion velocity. Moreover, optokinetic nystagmus is still present in monkeys where the conscious percept is diminished due to anesthesia with low doses of ketamine (Leopold et al., 2002).

Here, we introduce a new no-report paradigm that relies on active tracking of a fixation spot, unlike the reflex-based paradigms mentioned above. In this fixation-based paradigm, the subject is required to maintain fixation on a jumping spot, a task that many animals in vision research are already trained to perform. While following the fixation spot, subjects view either unambiguous, monocular stimuli physically switching between a face and an object, or a binocular rivalry stimulus that switches only perceptually. For the binocular rivalry stimulus, a fixation spot is shown to each eye at different positions on the screen. Thus, when the subject perceives a face in the left eye, he/she will generally perceive only the fixation spot in the left eye and saccade to it, ignoring the fixation spot in the right eye. In this way, the subject's percept can be inferred from his/her eye movement patterns without active report.

In a second innovation, we performed electrophysiological recordings using a novel 128-electrode site Neuropixels-like probe that allowed us to measure responses from large numbers of cells simultaneously. This allowed us to address for the first time the extent to which neural activity is modulated by conscious perception *in single trials*. Sheinberg and Logothetis (1997) found that 90% of IT cells were modulated by conscious perception, but the response modulations reported in that study during the rivalry condition were clearly smaller than those in the physical condition. This decrease could have been due to mixed selectivity of cells for the conscious percept and the physical stimulus on single trials. Alternatively, cells could have been modulated just as strongly by perceptual as by physical alternations and the decrease could have been due to incorrect reporting of the percept on some trials. Inter-trial averaging confounds these two possibilities.

To explore correlates of conscious perception, we targeted recordings to macaque face patches ML and AM. The macaque face patch system constitutes an anatomically connected network of regions in IT cortex dedicated to face processing (Tsao et al., 2006; Grimaldi et al., 2016; Chang and Tsao, 2017). To date, most response properties of cells in the face patch network can be explained in a feedforward framework without invoking conscious perception. For example, the functional hierarchy of this network, with increasing view invariance as one moves anterior from ML to AM (Freiwald and Tsao, 2010), can be explained by simple feedforward pooling mechanisms (Leibo et al., 2017). The representation of facial identity by cells in face patches through projection onto specific preferred axes can also be explained by feedforward mechanisms (Chang and Tsao, 2017). At the same time, it has been postulated that the fundamental architecture of the cortex may be a predictive loop, in which inference guided by internal priors plays a key role in determining what we see (Rao and Ballard, 1999). For example, one explanation for binocular rivalry is that it directly reflects our knowledge that two objects can't occupy the same space (Hohwy et al., 2008). The hierarchical organization of the face patch network, together with its specialization for a single visual form, makes it a promising testbed to examine the neural circuits' underlying construction of conscious visual experience, beyond feedforward filtering of visual input.

Here, we recorded from fMRI-identified face patches ML and AM in two monkeys using high-channel electrodes while we inferred the animals' conscious percept through the no-report paradigm described above. We found that high proportions of cells in both face patches (61% in ML and 81% in AM) encode the conscious percept even without active report. Population activity of perceptually-modulated cells was more weakly modulated during rivalry than during physical stimulus transitions in single trials. Nevertheless, we could still reliably decode the dynamically changing percept. Overall, these findings suggest that cells in macaque face patches encode both the physical stimulus and the animal's conscious visual percept.

Results. We first confirmed that it is possible to correctly infer a subject's conscious percept using a fixation-based no-report paradigm through a behavioral experiment in humans. We presented binocular rivalry stimuli consisting of a face (e.g., Obama) in the right eye and a non-face object (e.g., a taco) in the left eye, causing the percept to stochastically alternate between the two (Fig. VI-1a). Each of the stimuli contained a fixation spot that jumped to one of four possible locations every trial. Trials were

800 ms long and contained no blank period, i.e., stimuli were presented continuously. If subjects fixated at the fixation spot presented in the right eye on a given trial, we inferred that they perceived the face and vice versa for the object. To verify that the percept of face or object could be inferred from fixations, we instructed 6 naïve human subjects to perform the fixation task while simultaneously reporting their conscious percept with button presses. On trials where the percept switched, subjects also switched the fixation spot they were following (Fig. VI-1b). We were able to infer which image the subjects were consciously perceiving with accuracies ranging from 86% to 98% across subjects (average: 93%, Fig. VI-1c).

Fig. 1

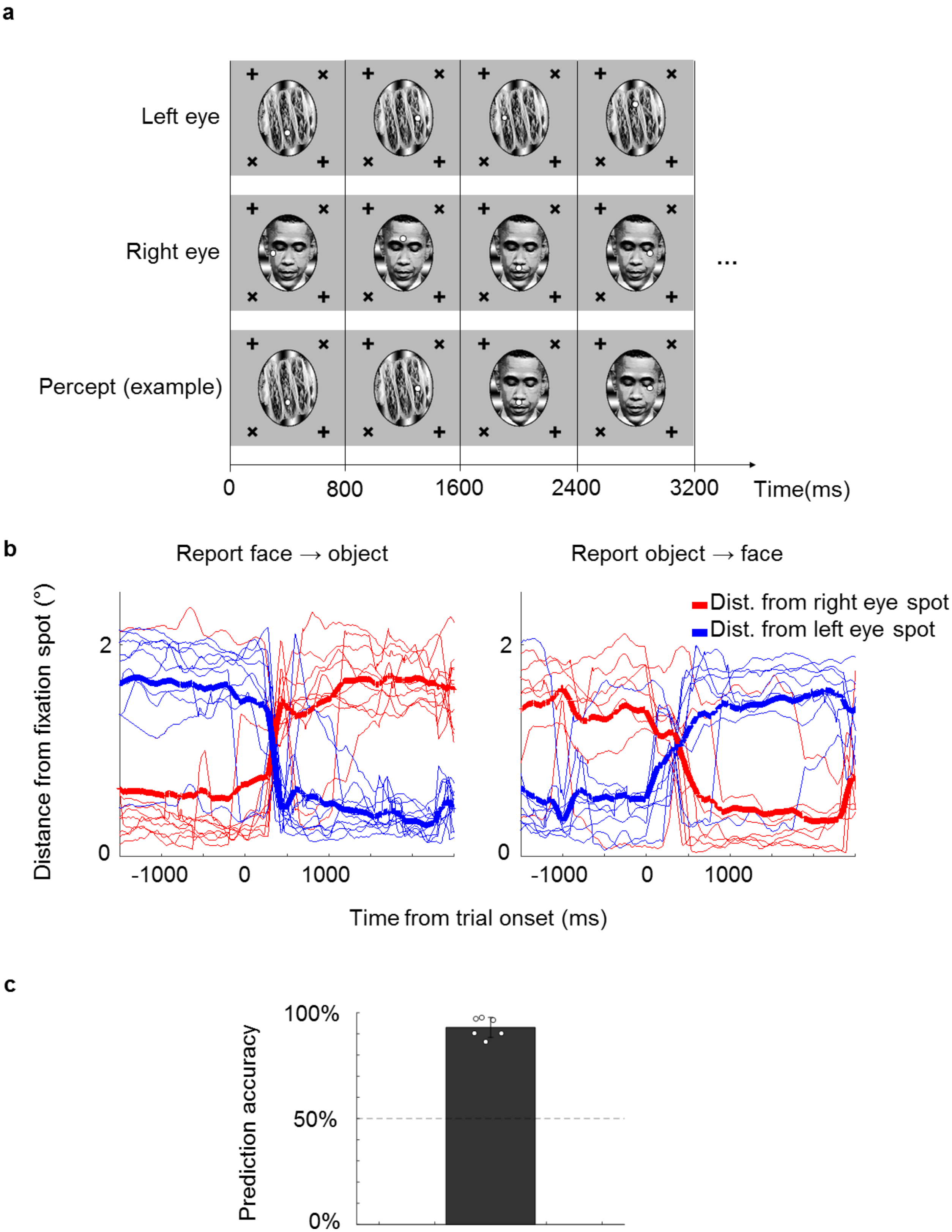

a
Left eye
Right eye
Percept (example)
...
0
800
1600
2400
3200
Time(ms)
b
Report face → object
Report object → face
Distance from fixation spot (°)
2
0
-1000
0
1000
2
0
-1000
0
1000
Dist. from right eye spot
Dist. from left eye spot
Time from trial onset (ms)
c
Prediction accuracy
100%
50%
0%

***Figure VI-1: A novel no-report paradigm. (a)** Illustration of binocular rivalry stimuli used in the paradigm. Four example trials are shown. Each trial was presented continuously for 800 ms each without blank period between trials. The first and second row show stimuli in the left and right eyes, respectively. If different stimuli are shown to the left and right eye, as in this example, one's percept will spontaneously alternate between the two, as shown in the example perceptual trajectory in the third row. Stimuli in each eye contained a fixation spot at one of four possible positions that the monkey was trained to fixate on. (**b**) Example eye traces from a human subject. Red and blue traces show the distance of the eye position from the fixation spot that is shown in the right and left eye, respectively. Thick lines show the average. Traces are aligned to the onset of a trial where the subject reported that the percept switched from face to object (left), or object to face (right). (**c**) The bar plot shows the average proportion of those trials where the percept inferred matched the percept reported by button press. White circles show accuracies of individual subjects. We inferred that a subject was perceiving face or object if the subject fixated on the face fixation spot (i.e., fixation spot in the eye of the face stimulus) or object fixation spot (i.e., fixation spot in the eye of the object stimulus), respectively, for at least half of the trial.*

We next used the same method in monkeys to infer their conscious percept while recording from face patches ML and AM in IT. Importantly, the two monkeys in this study had never been trained to report their percept. They had previously been trained to maintain fixation on a spot (presented binocularly) so they learned to perform the task within one or two days, respectively (reaching performance of maintaining fixation on a spot on at least 80% of all trials). We presented two types of stimuli. In the “physical” condition, unambiguous monocular stimuli were physically switched between face and object. In the “perceptual” (binocular rivalry) condition, the same face and object were continuously presented to the right and left eye, respectively, so any changes in percept were internally generated. To account for individuals’ eye dominance, we balanced the contrasts of the stimuli in the two eyes so that the monkey followed both fixation spots equally often in the rivalry condition. We inferred switches during rivalry when monkeys behaviorally switched from following the fixation spot in one eye to following the fixation spot in the other eye, as shown in the example eye traces in Fig. VI-2a, top. Spike rasters aligned to onset of trials where the percept switched from an example ML cell recorded in the same session are shown in Fig. VI-2a, bottom. Fig. VI-2b compares average response time courses to physical switches to face or object with responses to perceptual switches in example cells from ML and AM. Both example cells responded more strongly to a physically presented face than object, which is expected since they were recorded from face patches. Importantly, in the binocular rivalry condition, when the monkey perceived a face as inferred by its eye movement, the response of both cells was also higher than when the monkey perceived an object. Since the physical stimulus was identical in both cases, the response reflected its conscious percept of a face rather than just the physical input.

Fig. 2

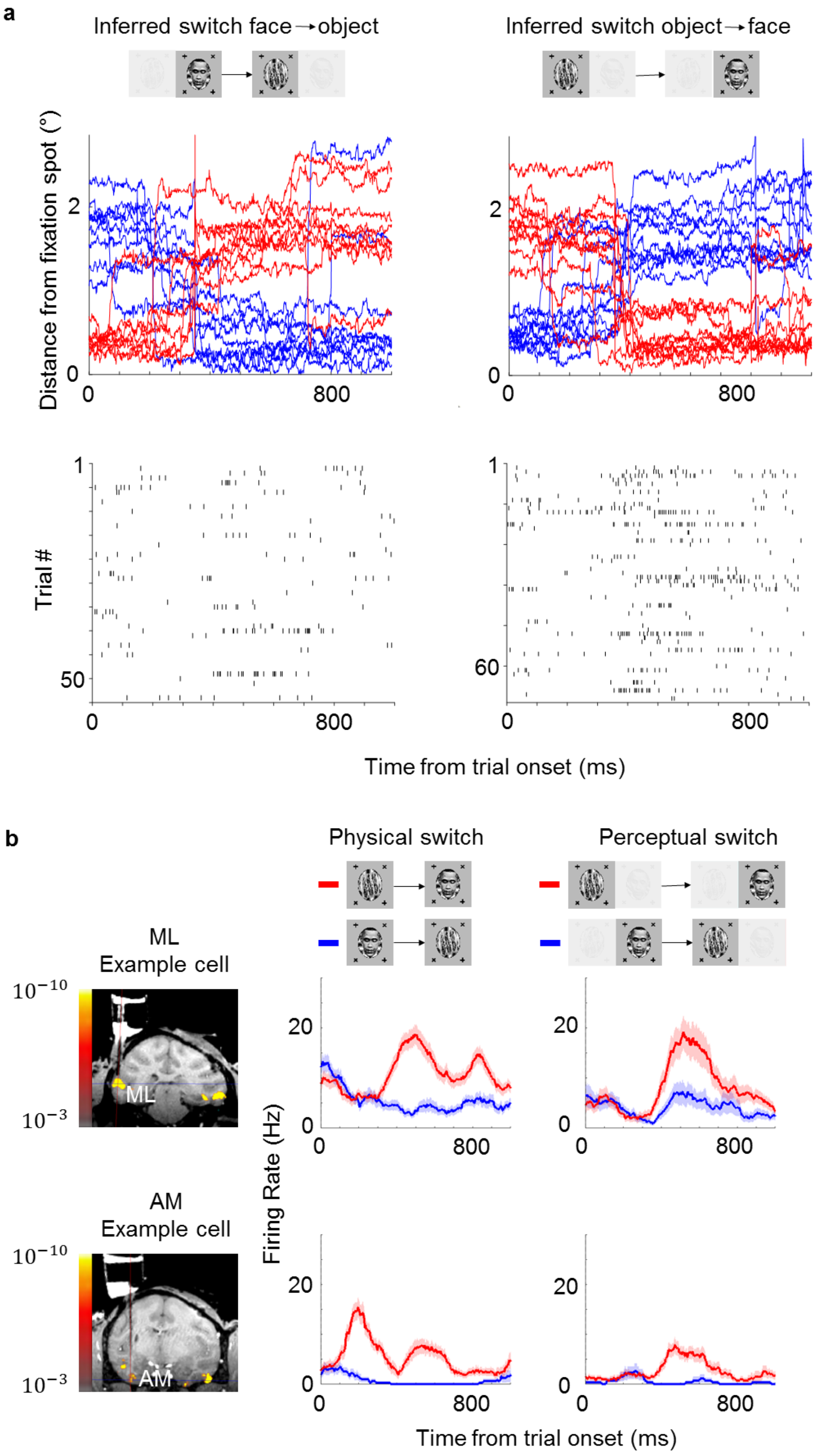

a
Inferred switch face → object
Inferred switch object → face
Distance from fixation spot (°)
2
0
0
800
2
0
0
800
1
Trial #
50
0
800
1
60
0
800
Time from trial onset (ms)
b
Physical switch
Perceptual switch
ML
Example cell
10^{-10}
10^{-3}
ML
AM
Example cell
10^{-10}
10^{-3}
AM
Firing Rate (Hz)
20
0
0
800
20
0
0
800
20
0
0
800
20
0
0
800
Time from trial onset (ms)

***Figure VI-2: Example face cells modulated by both physical and perceptual switches to face.**(**a**) Top: Example eye traces from a macaque performing the task aligned to a trial where the inferred percept switched from face to object (left) and from object to face (right), respectively. Red and blue curves indicate distances from the face and object fixation spots, respectively (as in Fig.* VI-*1b). Bottom: Spike raster of an example ML cell recorded in the same session as for the top panel. Responses are aligned to all trials where the inferred percept switched from face to object (left) and from object to face (right), respectively. (**b**) Left: Coronal slices from magnetic resonance imaging scan showing recording locations for the two example cells in this figure (top: face patch ML, bottom: face patch AM). Color overlay shows functional MRI activation to visually presented faces vs. non-face objects. Middle: Peristimulus histograms (PSTHs) show neuronal response time courses aligned to trial onsets where the visual stimulus was physically switched from face to object (blue) or from object to face (red). Right: PSTHs aligned to trial onsets where the inferred percept switched from face to object (blue) or object to face (red). ML cell is same cell as in (a). Shaded areas indicate standard error mean across trials.*

We recorded a total of 347 cells in ML and 210 cells in AM that were selective, i.e., showed a significant difference between face and object in the physical switch condition ($p<0.05$, two-sided t-test). Population results of all selective cells are shown in Fig. VI-3. Since we recorded from face patches, most cells showed stronger responses to the physically presented face stimulus. Importantly, most cells kept their preference in the perceptual condition. In face patch ML, 61% of cells were significantly modulated by the conscious percept in the binocular rivalry condition and showed preference consistent with the physical switch condition ($p<0.05$, two-sided t-test), while 9% of cells were significantly but inconsistently modulated. In AM, a face patch that receives input from ML (Grimaldi et al., 2016) and is the highest patch in the face patch hierarchy within IT (Freiwald and Tsao, 2010), the percentage of consistent modulation increased to 81%, with only 1% showing inconsistent modulation. For both patches, there was a clear correlation between modulation by physical stimuli and modulation by the percept in binocular rivalry ($r = 0.72, p < 10^{-31}$). Thus, in a no-report paradigm, cells in IT exhibit modulations by the conscious percept that reflect their response tuning to physically unambiguous inputs.

After eliminating the report confound, two important potential confounds remain. First, cells could be selective for the eye-of-origin of the fixation point that the animal is following (e.g., a cell could respond selectively to a fixation spot in the fovea of the left eye). Second, since we presented binocular stimuli using red-cyan anaglyph goggles, a confound could arise if cells were selective for the color of the fixation spot that is in the fovea. To control for these two potential confounds, we switched the colors and eye-of-origin of the face and object stimuli, i.e., where the face and its corresponding fixation spot were previously presented in red in one eye, they were now presented in cyan in the other eye and vice versa for the object (Fig. VI-3 supplement 1). If cells followed color or eye-of-origin, then all the dots in the upper right quadrant in Fig. VI-3, Supplement 1a should move to the lower left corner in Fig. VI-3, Supplement 1b. Instead, the majority of cells followed the object identity rather than color or eye-of-origin for both the physical and perceptual condition ($p < 10^{-29}$ for physical condition and $p < 10^{-11}$ for perceptual condition, one-sided t-test, alternative hypothesis that modulation indices are greater than 0). This confirms that cells in IT cortex indeed represent the conscious percept rather than the color or eye-of-origin of the fixation spot.

Fig. 3

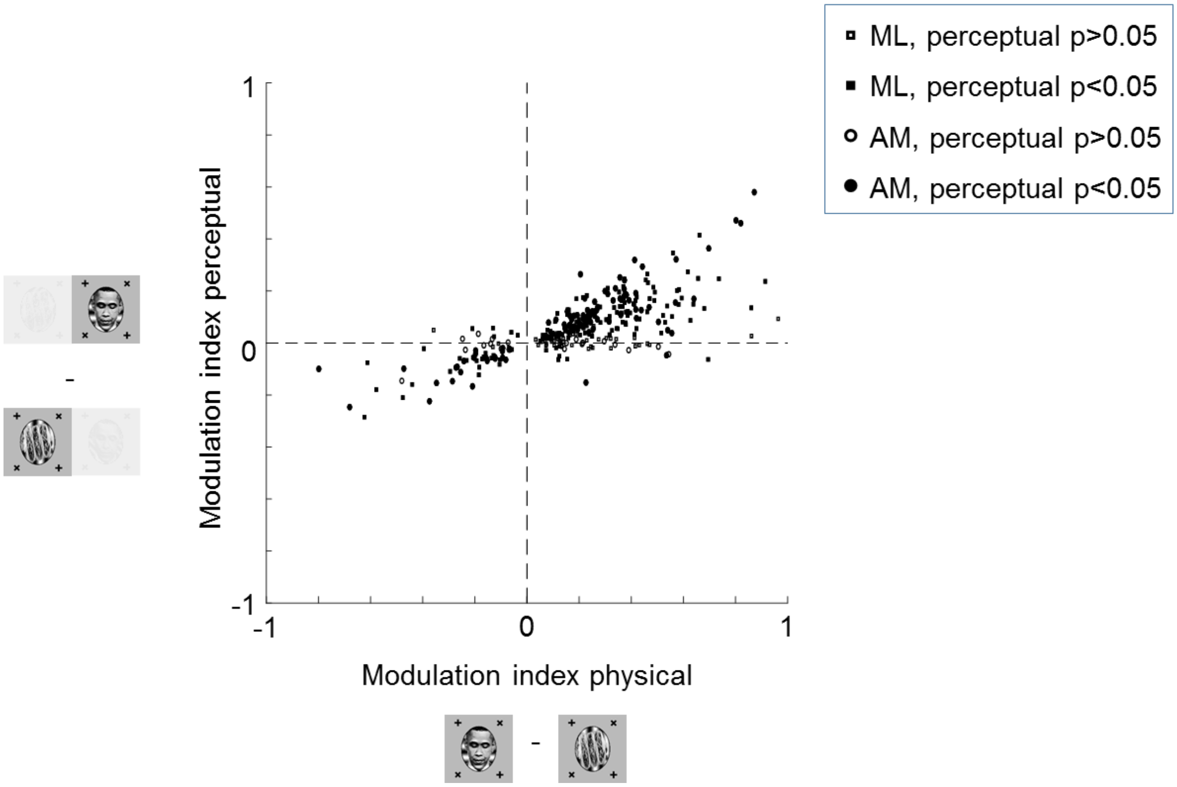


Figure VI-3: High proportions of face cells show modulation by conscious percept. *Scatter plot shows modulation indices* $\frac{R_{face}-R_{object}}{R_{face}+R_{object}}$ *measuring the difference in responses (i.e., average spike count* R*) on trials where the inferred percept was face or object, respectively, for the physical monocular condition (x-axis) and perceptual binocular rivalry condition (y-axis). Squares show cells from ML, and circles show cells from AM. Open and filled markers indicate cells without and with significant difference between perceived face and perceived object response in the binocular rivalry condition, respectively.*

To determine if one can decode the percept on a given trial from population activity, we performed recordings from multiple neurons simultaneously using S-probes with 32 electrode sites and passive Neuropixels-like probes with 128 electrode sites (see Methods for details). Fig. VI-4 shows recordings from face patch ML in one session using the Neuropixels probe. In this session, we recorded 81 cells simultaneously, of which 63 were face-selective (Fig. VI-4a). An example population time course snippet of cells recorded simultaneously in the perceptual switch condition showed clearly stronger activity across the recorded population during perception of face compared to object (Fig. VI-4b). The average population response across cells to perceptual switches is shown in Fig. VI-4c. We found above-chance decoding of the perceptual condition in all 12 sessions (in all but one session, responses were recorded in both ML and AM, and cells were pooled across the two patches). Cross-validated accuracies of linear classifiers across different sessions are shown in Fig. VI-4d (see Methods). Decoding accuracies were 99% for the best session and 95% on average for the physical condition. For the perceptual condition, decoding was 88% on the best session and 78% on average.

Fig. 4

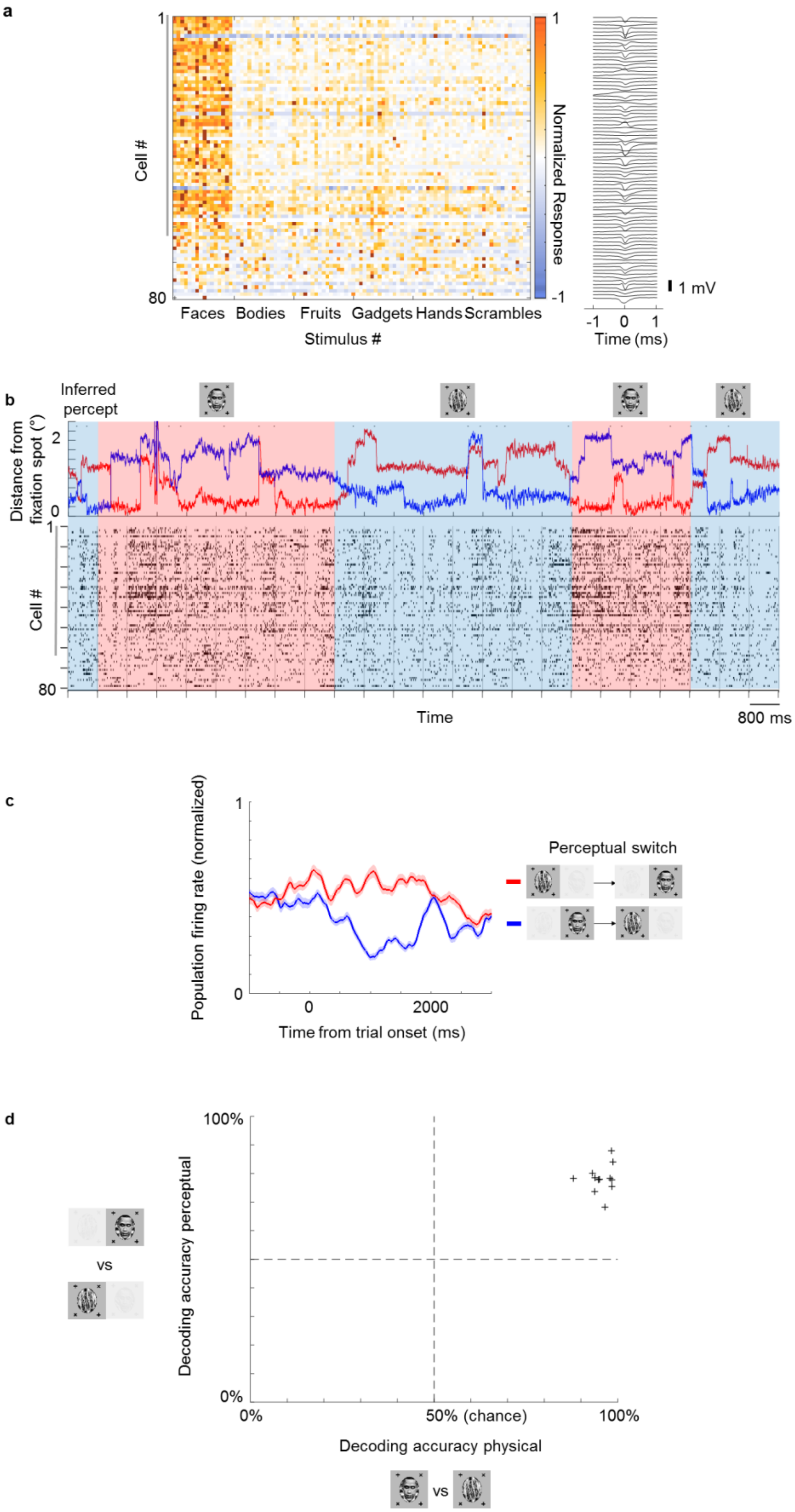

a
1
80
Cell #
Faces Bodies Fruits Gadgets Hands Scrambles
Stimulus #
1
-1
Normalized Response
1 mV
-1 0 1
Time (ms)
b
Inferred percept
Distance from fixation spot (°)
2
0
1
Cell #
80
Time
800 ms
c
Population firing rate (normalized)
1
0
0
2000
Time from trial onset (ms)
Perceptual switch
d
100%
0%
Decoding accuracy perceptual
vs
0%
50% (chance)
100%
Decoding accuracy physical
vs

Figure VI-4: Multi-channel recordings allow decoding of conscious percept on single trials. *(**a**) Left: Average responses (baseline-subtracted and normalized) of cells (rows) to 96 stimuli (columns) from 6 categories, including faces and other objects. Right: Waveforms of cells corresponding to rows on the left. Face-selective cells indicated by gray vertical bar on left. (**b**) Top: Example eye trace across 24 trials as in Fig. VI-1b in a binocular rivalry session (i.e,. only perceptual, no physical switches). The inferred percept across trials according to eye trace is indicated by shading (red = face, blue = non-face object). Small black dots on top of eye traces indicate time points where our method detected saccades (see Methods), which were used in Fig. VI-5 and Fig. VI-5, Supplement 1. Bottom: Response time course snippet of a population of 81 neurons recorded with a Neuropixels probe in ML simultaneously to the eye trace at top. Each row represents one cell; ordering same as in (a). Face-selective cells indicated by gray vertical bar on left. (**c**) Normalized average population response across all significantly face-selective ML cells recorded from one Neuropixels session (same session as in (a), (b) to perceptual switch from object to face (red) and face to object (blue). Shaded areas indicate standard error mean across cells. (**d**) Cross-validated decoding accuracy of a linear classifier trained to discriminate trials of inferred percept face vs. inferred percept object for the physical switch condition (x-axis) and perceptual switch condition (y-axis). Each plus symbol represents a session of neurons recorded simultaneously with multi-channel electrodes.*

Looking at the population time course, we noticed bursts of activity that appeared to be triggered by saccades, which occurred even when an object was perceived (blue epochs in Fig. VI-4b; small black dots on top indicate detected saccades). This raised the possibility that cells modulated by perception may still carry information about the physical stimulus. To investigate this further, we selected cells that (1) showed both significant physical and perceptual modulation and (2) consistently preferred the face over the object. We then averaged responses across these cells and computed response time courses triggered by individual saccades, grouped by whether a saccade occurred during a trial inferred to be face or object, respectively (Fig. VI-5). We observed response modulations for both physical and perceptual conditions starting around 130 ms after saccade onset (Fig. VI-5a). In the physical condition, a saccade during an object epoch led to response suppression, while a saccade during a face epoch led to response increase. In striking contrast, in the rivalry condition, saccades led to response increase in both object and face epochs. As a consequence, during rivalry the response difference to a saccade between face and object, though significant ($p = 10^{-23}$, two-sample t-test), was weaker than during the physical condition. Computing histograms of responses averaged across neurons for individual saccades shows that responses in the rivalry condition were less bimodal and spanned a smaller range compared to the physical condition (Fig. VI-5b). Importantly, this difference in response profiles between physical and perceptual conditions was apparent even when pooling across both face and object trials (Fig. VI-5b, middle), and *hence cannot be explained by mistakes in inferring the percept from eye movements*. We computed the absolute value of these responses and found the difference in response distributions to be significant (Figure VI-5b, right, $p = 6 \cdot 10^{-35}$, two-sample t-test on absolute value distributions).

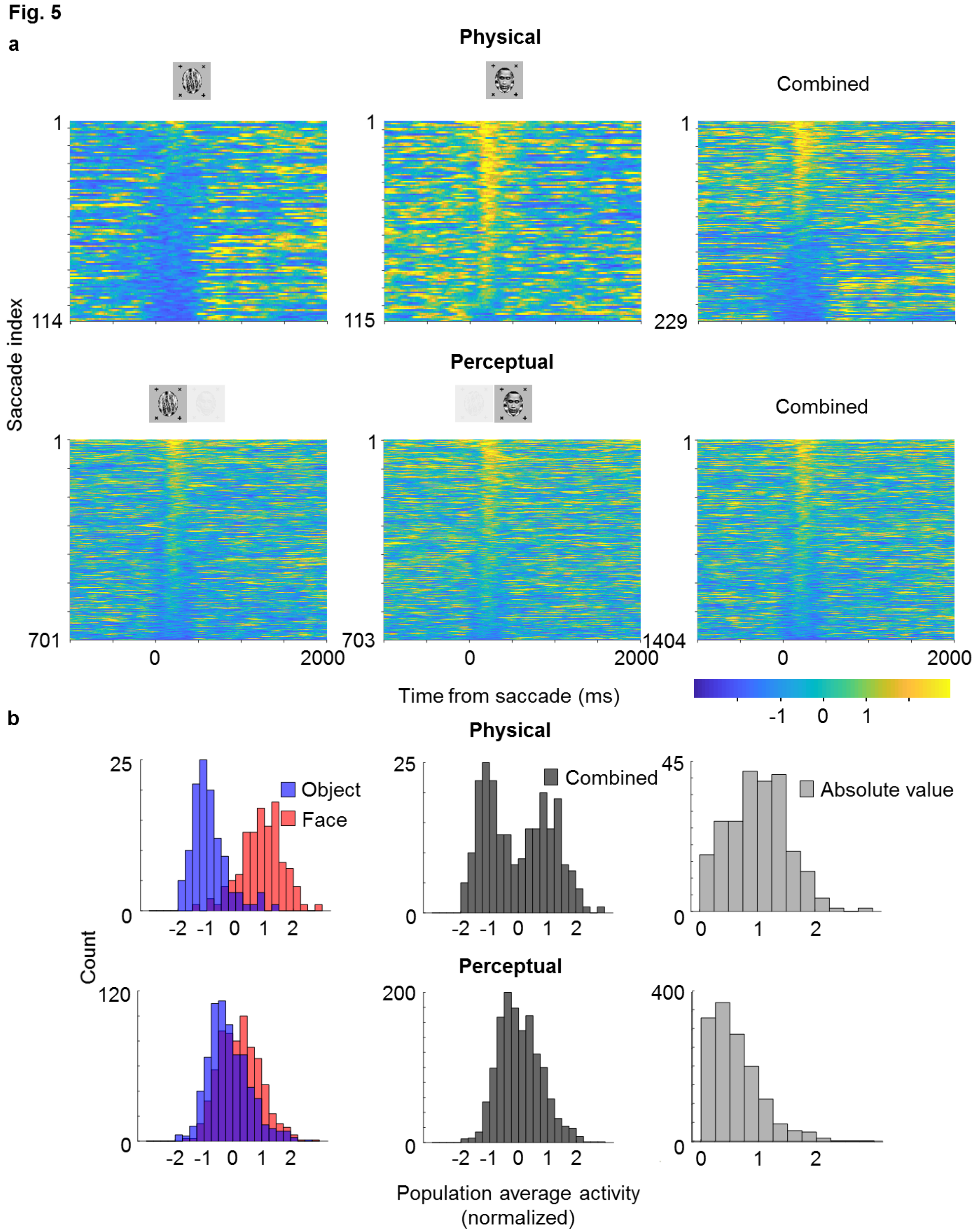


***Figure VI-5: Saccade-triggered responses are less bimodal during rivalry.** (a) Single-trial responses during saccades averaged across simultaneously recorded*

ML neurons from the same session as in Fig. VI-*4b that were significantly face-selective for both physical and perceptual condition. Individual neuron responses were normalized to make -1 correspond to mean object response, 1 correspond to mean face response, and 0 correspond to the average of the two. Rows of each plot correspond to response time courses to individual saccades, aligned to saccade onset, and sorted by average response during 0 to 400 ms after saccade onset. Top: Physical condition. Bottom: Perceptual condition. Left, middle, and right columns correspond to saccades during (inferred) object, face, and across both, respectively. The difference between perceptual and physical conditions in the third column shows that this difference cannot be simply attributed to mislabeling of perceptual state by the no-report paradigm. (**b**) Histograms of saccade-aligned responses averaged across a time window of 0 to 400 ms after saccade onset and across neurons (after normalizing as in (a)) that were significantly modulated for both physical and perceptual condition. Blue, red, and gray responses correspond to counts of saccade responses during object, face, and either, respectively. Top: Physical condition. Bottom: Perceptual condition. Left: Saccades for face and object plotted separately in red and blue, respectively. Responses were normalized to be 0 if the response was equal to the average of the face and object response, and 1 if equal to either the average face or average object response. Middle: Saccades for either face or object plotted in grey. Right: Absolute values of normalized responses plotted in light grey.*

The observation of different response profiles for physical and perceptual conditions was not specific for saccades: histograms were also less bimodal and spanned a smaller range for the rivalry condition when triggering responses on trial onsets rather than saccades in both ML (Fig. VI-5, Supplement 1a, $p = 9 \cdot 10^{-15}$) and AM (Fig. VI-5, Supplement 1b, $p = 0.0014$). Therefore, it appears that throughout rivalry, for perceptually-modulated cells, response differences between face and object are less pronounced than in the physical condition, and this is true in both ML and AM. One tantalizing explanation for this phenomenon is that perceptually-modulated cells may be multiplexing information about both the physical stimulus and the perceptual state during single trials, allowing both to be simultaneously represented across the face patch hierarchy.

Discussion. We have shown that face patches ML and AM in macaque IT cortex are modulated by conscious perception and do not merely encode the physical input. Importantly, monkeys in this study had never been trained to actively report their percept. Instead, we were able to infer their percept from eye movements using a new no-report paradigm. Thus, activity modulations attributed to switches in conscious perception in IT cannot be explained simply by active report.

Previous single-unit recordings in IT cortex using active report to infer the percept found 90% of cells represent the conscious percept (Sheinberg and Logothetis, 1997). Here, we found proportions of 61% in ML and 81% in the more anterior patch AM. The quantitative difference may be due to several factors including different recording sites (Sheinberg and Logothetis recorded from both upper and lower banks of the superior temporal sulcus in a less specifically targeted manner), imperfect accuracy of the no-report paradigm, and differences in stimuli and analysis methods. Importantly, our results show that the majority of cells in IT cortex do represent conscious perception and not merely active report and its accompanying cognitive factors. Furthermore, this new paradigm makes studies of consciousness in monkeys more accessible, by replacing the need to train the animal to signal its conscious percept (which can be a laborious process) with a simple task that only requires animals to follow a fixation spot.

Our results show that for cells that are modulated by conscious perception, the modulation is not “all-or-none.” First, we found that the average response modulation during the perceptual condition was weaker than during the physical condition (Fig. VI-3). This was also found in a previous study of rivalry (Sheinberg and Logothetis, 1997). This could be explained either by incomplete modulation, or by imperfect labeling of the animal’s perceptual state. The key question is: *what* *happens during single trials?* In the rivalry condition, do responses in single trials look like those to either physically-presented faces or objects? By recording from a large number of face cells simultaneously using a novel 128-electrode site probe specifically designed for use in primates, we could address this question for the first time. Surprisingly, we found a dramatically different response profile on single trials between the perceptual and physical conditions (Fig. VI-5). Whereas in the physical condition responses clustered into two groups, in the rivalry condition, responses appeared unimodal, lying in between the two clusters for the physical condition. This suggests that single cells are multiplexing both the conscious percept and the veridical physical stimulus during single trials, such that information about both the perceived and unperceived stimuli remain constantly available in IT cortex. Future experiments varying the identity of the unperceived stimulus will be needed to further test this hypothesis. An alternative explanation is that cells are not modulated by the identity of the suppressed stimulus, and simply encode the dominant stimulus with reduced gain when presented in rivalry.

Compared to previous approaches that attempted to isolate representations of the conscious percept, our new no-report binocular rivalry paradigm has several advantages: For flash suppression, where a stimulus flashed in one eye suppresses the stimulus in the other eye, report is also not required (Wolfe, 1984; Wilke et al., 2003; Tsuchiya and Koch, 2005). However, in that case, the physical input when the target is perceived vs. when it is suppressed is not identical, and thus any modulation observed may be driven entirely externally. Indeed, it is known that if a distractor stimulus is presented simultaneously with a preferred stimulus, the response can be reduced compared to when the preferred stimulus is presented alone as a result of simple normalization mechanisms (Bao and Tsao, 2018). Another paradigm that has been widely used to study the neural correlates of consciousness is backward masking. Here, the stimulus is presented for such a short time before being masked that sometimes it enters consciousness and sometimes not (Breitmeyer et al., 1984). So far, backward masking has always relied on report. Also, it is more susceptible to modulations arising from bottom-up withdrawal of attention or low-level (e.g., retinal) noise, whereas in binocular rivalry, perceptual switches appear to be internally generated. One potential confound described by Block as the "bored monkey problem" is that the monkey may still be thinking about whether it is perceiving object or face and internally report it even if it is not required to actively report it (Block, 2020). It is methodologically very difficult to entirely remove this confound, but the fact that monkeys had to simultaneously perform a very challenging unrelated task of saccading to jumping fixation points should at least alleviate this concern. Thus, to the best of our knowledge, this study shows representations of the conscious percept in IT cortex in the most confound-free way to date. Our study complements a study conducted in parallel by Kapoor et al. (2020) that found modulations by conscious percept in prefrontal cortex using a different no-report paradigm based on optokinetic nystagmus.

The existence of two directly-connected functional modules with a hierarchical relationship (ML, AM) that both encode the conscious percept of a particular type of object opens up the possibility for future studies to investigate how changes in the conscious percept are coordinated across the brain. Recordings and perturbations in multiple face patches simultaneously using high-channel population recordings may reveal the dynamics of information flow, e.g., whether switches occur in a feedforward or feedback wave. This may yield insight into the mechanism for how a conscious percept emerges in the brain as an interpretation of the world that is consistent across different levels of representation.

Methods. All animal procedures in this study complied with local and National Institute of Health guidelines including the US National Institutes of Health Guide for Care and Use of Laboratory Animals. All experiments were performed with the approval of the Caltech Institutional Animal Care and Use Committee (IACUC). The behavioral experiment with human subjects for the human psychophysics experiment complied with a protocol approved by the Caltech Institutional Review Board (IRB).

Targeting. Two male rhesus macaques were implanted with head posts and trained to fixate on a dot for juice reward. We targeted face patches ML and AM in IT cortex for electrophysiological recordings. ML and AM were identified using functional magnetic resonance imaging (fMRI). Monkeys were scanned in a 3T scanner (Siemens), as described previously (Tsao et al., 2006). MION contrast agent was injected to increase signal-to-noise ratio. During fMRI, monkeys passively viewed blocks of faces and blocks of other objects to identify face-selective patches in the brain. Recording chambers (Crist) were implanted over ML and AM. Guide tubes were inserted into the brain 4 mm past the dura through custom printed grids placed inside the chamber, and electrodes were advanced to the target through the guide tube. Both chamber placement and grid design were planned with the software Planner (Ohayon and Tsao, 2012). After insertion of tungsten electrodes, correct targeting of the desired location was confirmed with anatomical MRI scans.

Electrophysiology. Recordings were performed using tungsten electrodes (FHC) with 1 MΩ impedance and, after correct targeting was confirmed, with 32-channel S-probes (Plexon) with 75 µm and 100 µm inter-electrode distance, and with passive Neuropixels-like probe prototypes (IMEC) (Jun et al., 2017; Dutta et al., 2019). These prototypes were a limited stock of test devices that were developed and used for testing as part of the development of primate Neuropixels probes and are not available for other labs. Unlike the final product, the prototypes had 128 passive electrode sites across 2 mm (arranged in two parallel staggered bands), but used the same electrode materials and shank specifications (45 mm total shank length). All electrodes were advanced to the target using an oil hydraulic Microdrive (Narishige). Neural signals were recorded using an Omniplex system (Plexon). Local field potentials were low-pass filtered at 200 Hz and recorded at 1000 Hz, and units were high-pass filtered at 300 Hz and recorded at 40 kHz. Only well-isolated units were considered for further analysis.

Task. Monkeys were head fixed and viewed an LCD screen (Acer) of 47-degree size in a dark room. Monkeys viewed stimuli of 5-degree size wearing red-cyan anaglyph goggles custom made with filters to match the red and green/blue emission spectrum of the screen, respectively, so that inputs to left and right eye could be controlled independently. Emission spectra were measured using a PR-650 SpectraScan colorimeter (Photo Research). Eye position was monitored using an eye tracking system (ISCAN). In the first phase of the experiment, monkeys passively viewed at least 5 repeats of 61 screening stimuli in pseudorandom order (250 ms ON time, 100 ms OFF time) with a fixation spot of 0.25 degree diameter in the center of the screen. Screening stimuli consisted of 20 images of faces and 41 images of non-face objects. During this phase, monkeys received a juice reward for maintaining fixation for at least 3 seconds. Subsequently, for the main experiment, stimuli contained one or two fixation spots at one of four possible locations (top, bottom, left, and right, 1 degree from the center) and were presented for 800 ms ON time and 0 ms OFF time. In the case of two fixation spots, stimuli contained one fixation spot per eye, and the two spots never appeared at the same location. During this phase, the monkey received a juice reward if it maintained fixation within 0.5 degree of one of the fixation spot for at least half of the trial duration (i.e., 400 ms, not required to be contiguous). Stimuli during the main experiment included (1) a monocular face/monocular object with one fixation spot, and (2) a binocular stimulus composed of a face and a fixation spot in one eye, and an object and a second fixation spot in the other eye. To improve rivalry and reduce periods of mixture, face and object stimuli were presented on backgrounds consisting of gratings that were orthogonal in the two eyes. Moreover, we applied orthogonal orientation filters (with concentration $\sigma_{angle} = 0.5°$) to the face and object stimuli, respectively, to increase local orientation contrast.

Online analysis. Spikes were isolated and sorted online using the PlexControl software (Plexon). During the screening phase, the average number of spikes during the time window from 100 ms to 300 ms was calculated for each unit and stimulus. For each stimulus, the average response across units was determined after normalizing the response of each unit by subtracting the mean and dividing by the standard deviation for the unit. Subsequently, the face stimulus with the highest average response and the object stimulus with the lowest average response were chosen to generate stimuli for the main experiment.

Offline analysis. For human subjects, the inferred percept based on button-presses on a given trial was determined according to the last report the subject made before the end of the trial. For humans and monkeys, we also determined their inferred percept based on eye movements depending on which fixation spot they fixated on if they fixated on one of the fixation spots for at least half of the trial duration (i.e., 400 ms, not required to be contiguous). We computed L-1 norms for computing the distance between eye position and a given fixation spot. We accounted for a saccade delay of on average 350 ms, by analyzing the eye data 350 ms until 1150 ms after trial onset. For Figures VI-3 and supplement, VI-4d and VI-5 supplement, in order to exclude trials during which the percept switched back to the opposite percept, we also required the following trial to have the same inferred percept as the current trial. Spikes were re-sorted using the software OfflineSorter (Plexon). For Neuropixels, since the high density of electrodes allowed the same neuron to appear on multiple channels, we used Kilosort2 to re-sort spikes (Pachitariu et al., 2016). A total of 551 and 408 cells were recorded in monkey A and monkey O, respectively. To correct for delays in stimulus presentation, we used a photodiode that detected the onset and offset of the stimuli. The output of the photodiode was fed into the recording system and later used to synchronize the onset of the stimulus and the neurophysiological data during offline analysis. Peristimulus time histograms (PSTHs) were smoothed with a box kernel (100 ms width). For computing modulation indices, we used the average spike count across trials as response.

Decoding analysis was performed with a support vector machine with a linear kernel (Matlab fitcsvm) trained to discriminate trials where the inferred percept was face or object, respectively. As predictor variables, we used the spike count during the 800 ms of each trial for all simultaneously recorded neurons. All decoding accuracies were cross-validated (leave-one-out). In more detail, one trial was chosen for testing and the rest of the trials for training, and this was repeated for all trials to compute decoding accuracies. Criteria for detecting a saccade were as follows. A saccade was detected at time t if the distance between the mean eye position during t-100,...t-2 ms and the mean eye position during t+2,...t+100 ms was greater than 0.5 degree, and the eye position during t-100,...t-2 ms and t+2,...t+100 ms, respectively, stayed within 0.5 degree of the respective mean for at least 80% of the duration of each period. We also required consecutive saccades to be at least 100 ms apart from each other. All analysis was performed using Matlab (MathWorks).

Acknowledgements

This work was supported by HHMI and the Simons Foundation. We are grateful to members of the Tsao lab for feedback on the manuscript, Varun Wadia for helping us collect the human subject data, Audo Flores for animal support, Daniel Wagenaar and Eric Trautmann for technical assistance, and Barun Dutta, Tim Harris, Tirin Moore, Michael Shadlen, Krishna Shenoy, and HHMI for contributions to development of NHP Neuropixels probes.

Supplementary material

Fig. 3 supplement 1

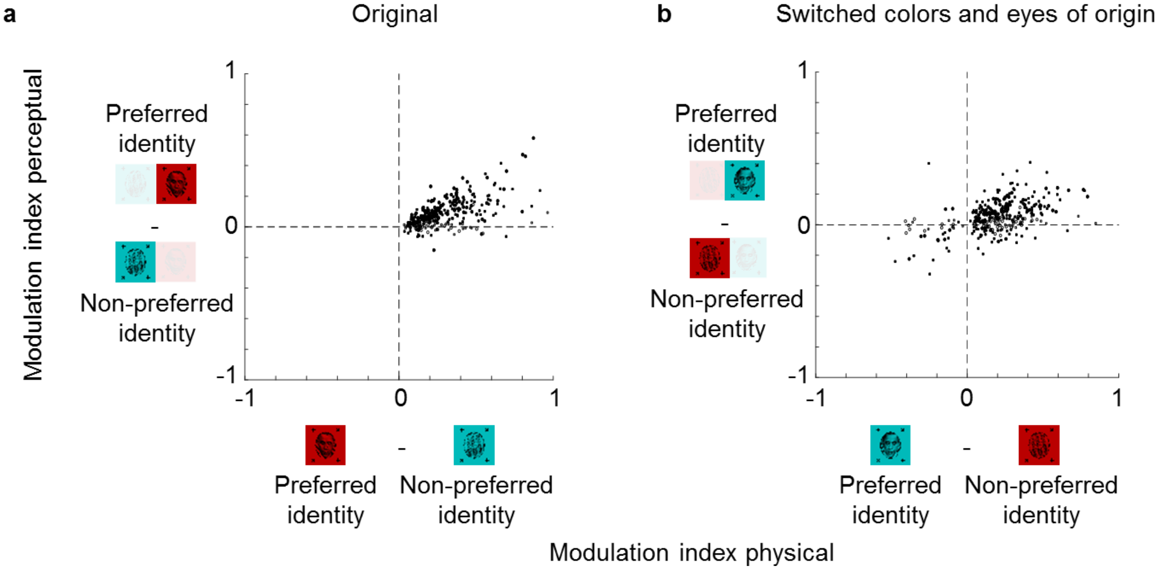


Figure VI-3, Supplement 1: Color and eye-of-origin confound control.

Left: Scatter plot similar to Fig. VI-3*, but modulation indices* $\frac{R_{preferred}-R_{nonpreferred}}{R_{preferred}+R_{nonpreferred}}$ *now show the difference between preferred and non-preferred stimulus. The preferred stimulus is face if the response to face is higher and non-face object if the response to non-face object is higher in the physical condition. Thus, by definition the x-values of all cells are positive. Right: Scatter plot of modulation indices* $\frac{R_{preferred}-R_{nonpreferred}}{R_{preferred}+R_{nonpreferred}}$ *for the same preferred and non-preferred object identities of stimuli when the colors and eye of origin of the two stimuli were switched; importantly, the preference of a given stimulus identity was assigned based on responses to stimuli of the original color and eye of origin stimulus responses. N = 192 for ML and N = 120 for AM for both plots.*

Fig. 5 Supplement 1

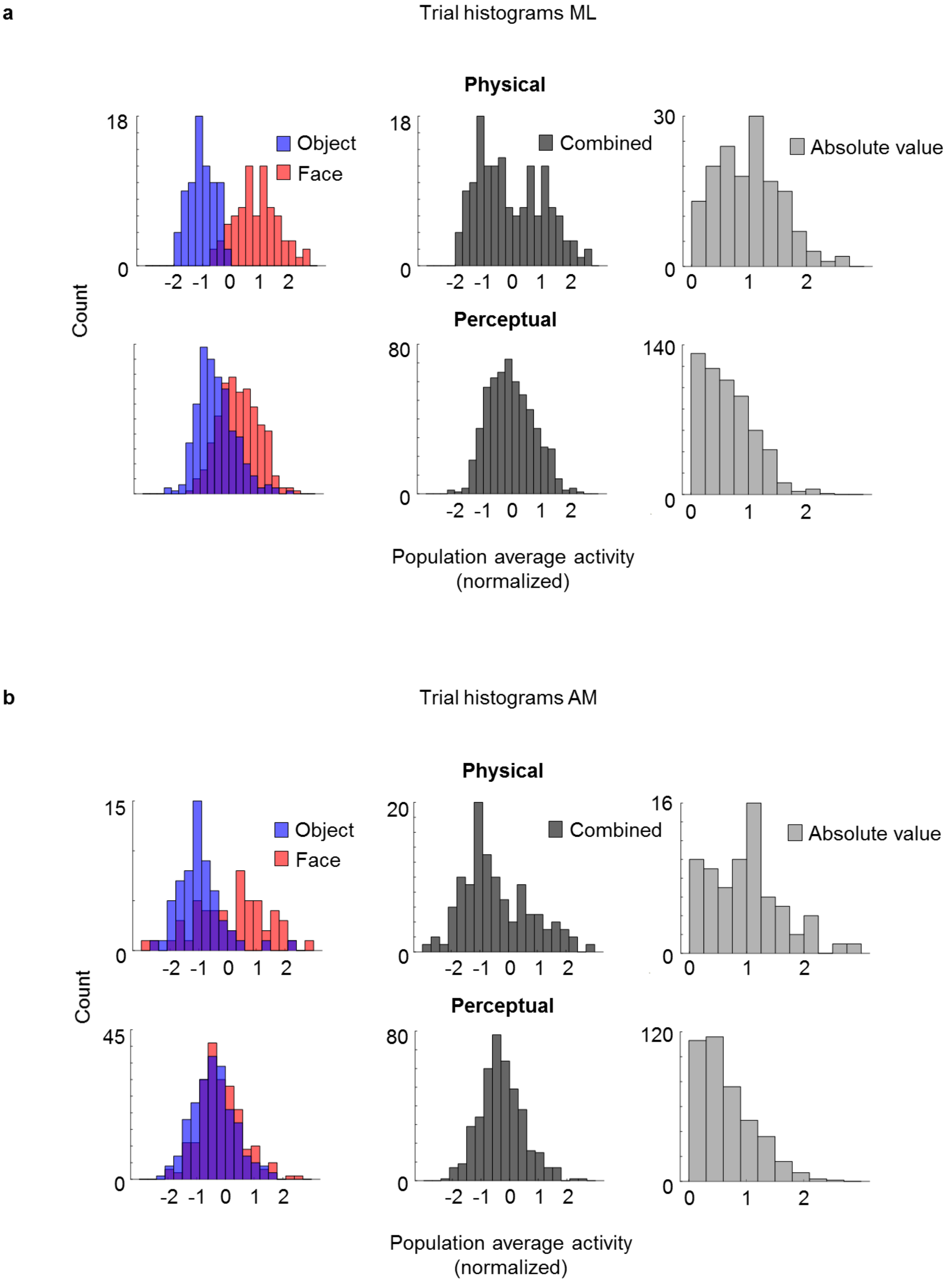


Figure VI-5, Supplement 1: Lack of bimodality is a general trademark of rivalry. *(**a**) Trial responses in ML are less bimodal during rivalry. Histograms have same conventions as Fig.* VI-*5b, but instead of averaging neuron responses for individual saccades, responses are averaged across trial duration for individual trials. (**b**) Trial responses in AM are less bimodal during rivalry. Same conventions as in (a), but instead of the Neuropixels-like probe in ML, cells were simultaneously recorded from AM. Due to technical limitations, the 128-channels Neuropixels-like probe did not reach the depth of AM, and cells were recorded using a 32-channel S-probe instead.*

Recordings in human epilepsy patients

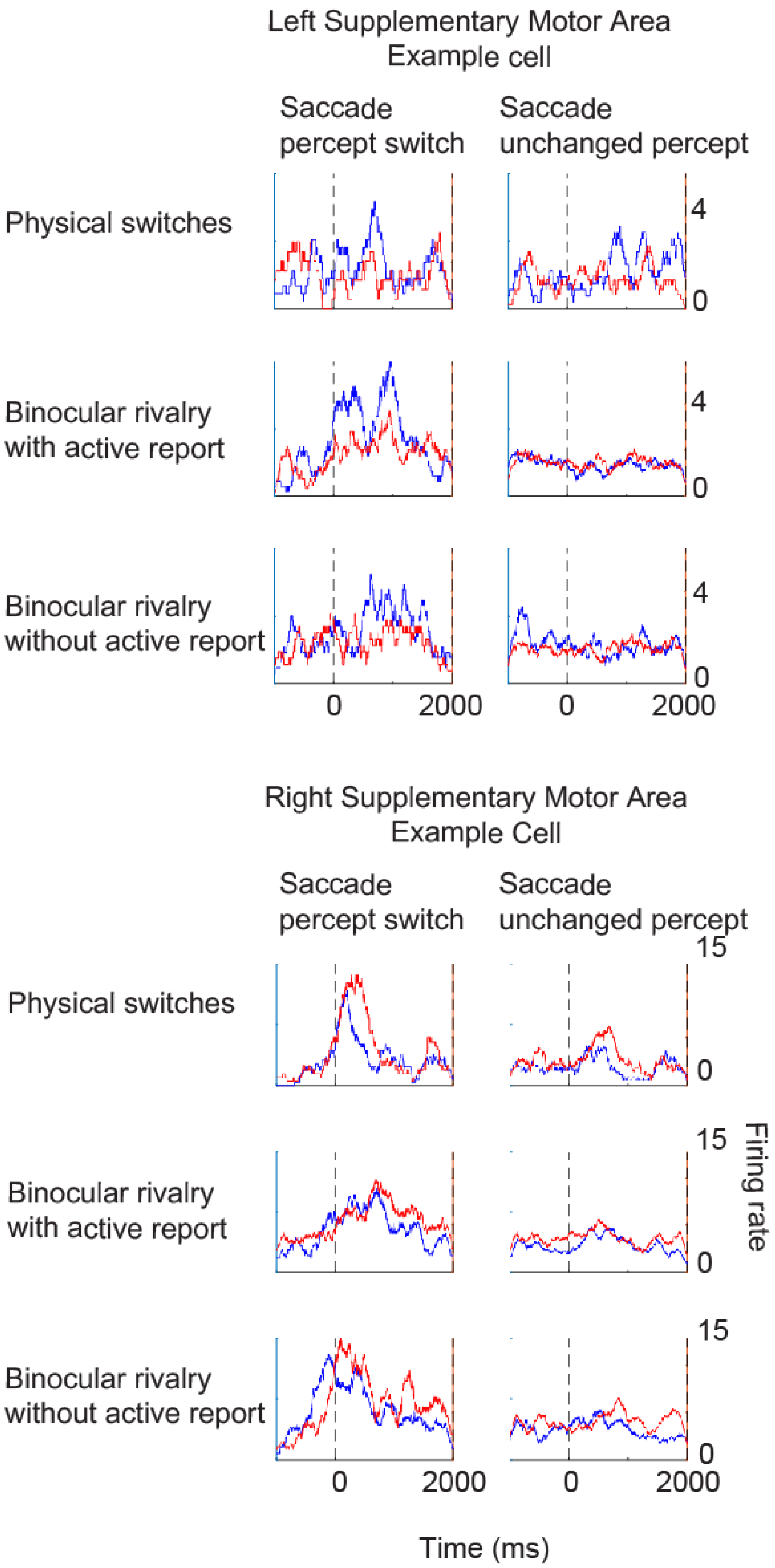
Besides studying binocular rivalry without report in macaque face patches, we have also conducted preliminary experiments in humans. This work was done in collaboration with Varun Wadia and Ueli Rutishauser at Cedars Sinai in epilepsy patients that had been implanted with depth electrodes for surgical treatment of epilepsy. The advantage of this system is that, in humans, we can directly compare modulations during active report and without report since humans can be easily instructed to report their percept, as compared to a lengthy training process in animals. Moreover, the human patients had electrodes implanted across a large number of brain areas (including pre-supplementary motor area, amygdala, orbitofrontal cortex, anterior cingulate cortex, and hippocampus), in order to localize the origin of seizures, allowing us to investigate representations of conscious percept across all these areas simultaneously and holding the potential to study interactions between them.

We used the same paradigm for inferring conscious percept from eye movement as in monkeys. Unless stated otherwise below, methods were the same as in the previous section of this chapter. In addition to following fixation spots, subjects were at one part of the experiment also instructed to report their percept by pressing one of two buttons (indicating perceived stimulus 1 and perceived stimulus 2, respectively) on a response pad (Cedrus). All four subjects in this preliminary study volunteered for the study and gave informed consent. This study was approved by the Institutional Review Boards of Cedars-Sinai Medical Center and Huntington Memorial Hospital. Instead of ISCAN, we used an EyeLink camera (SR Research) to track eye position. We recorded from up to ten surgically implanted macroelectrodes, each containing eight 40-μm diameter microwires (Rutishauser et al., 2010; Minxha et al., 2018). We recorded broadband (0.1-9000Hz filter) signal sampled at 32 kHz using a Neuralynx Atlas system (Neuralynx). Signals were locally referenced to one of the eight microwires in a given brain area. For spike-sorting, we used a semi-automatic template matching algorithm described in (Rutishauser et al., 2006). Only well-isolated units were included in further analysis.

In an initial session, usually in the morning of the same day as the main experiment, we let the patient passively view a set of 63 screening stimuli. Screening stimuli were grayscale images of faces and objects and were presented for 10 repetitions each with an ON time of 500 ms and a blank OFF period varied between randomly between 600 and 900 ms. Subjects were instructed to passively view the screening stimuli, however, at random times every 20-40 images they were presented with a yes/no catch question about the previous image that subjects answered by button press, in order to have subjects keep paying attention to viewed stimuli. We then selected two stimuli that elicited the strongest responses in two (non-overlapping) sets of neurons, and generated binocular rivalry stimuli from them.

After analyzing the results of the screening session and generating stimuli, we performed the main experiment for each patient. The main experiment consisted of several stages, including a (1) calibration stage to adjust the contrast of the two binocular rivalry stimuli until the subject followed the two fixation spots equally often, (2) a physical switches stage, where the viewed stimulus switched physically between two monocular stimuli, while subjects followed the one fixation spot, (3) a binocular rivalry without report stage, where subjects perceived rivalry switches following one of the two fixation spots, and (4) a binocular rivalry stage with report stage where subjects followed one of the two fixation spots and at the same time actively reported perceptual switches by button press. The only difference between the stimuli presented to humans and stimuli presented to monkeys was that for humans, fixation spots changed position every 2000 ms rather than 800 ms to allow for more time to find the fixation spot, given the lack of training compared to monkeys.

Doing the task in humans, we were faced with some challenges, such as getting good eye signals through corrective glasses plus red-cyan filters and finding cells responding strongly to one of the stimuli during stimuli that persisted until the afternoon session, hence the data presented here is preliminary. Yet, from the experiment stage where we both inferred the percept from eye movement and instructed subjects to report their percept, we were able to infer that the paradigm worked well to predict their button presses from eye movement, with average accuracies of 87%. Below are two example cells from left and right pre-SMA (Fig.VI-6). The two cells showed responses to a physical switch from stimulus 1 to stimulus 2 (red) or from stimulus 2 to stimulus 1 (blue), with a slight preference for switches to stimulus 1 (top cell), or switches to stimulus 2 (bottom cell), respectively. Importantly, both cells showed responses even in the binocular rivalry condition without any physical change, when a switch of percept was inferred from eye movements, and appeared to keep their preference. This was true whether the patient actively reported their percept or not, showing that this pre-SMA activity was not just due to report and its accompanying factors. Given that the selectivity was not very strong, it is important to also show that the apparent response to perceptual switches was not just due to saccades and accompanying afferent copies of motor responses or changes in visual input. The right column thus shows activity aligned to saccades where the percept remained constant, e.g., where the patient was perceiving stimulus 1, and made a saccade because the fixation spot in the left eye changed position but kept following the fixation spot in the left eye (blue). Responses aligned to these events were relatively flat. This suggests that responses in pre-SMA encode switches in conscious percept rather than confounding factors such as report or saccades. Note that this data is still preliminary and further experiments with more selective cells are needed to confirm these findings and determine coding of conscious percept in the other brain areas that were recording targets.



Left Supplementary Motor Area
Example cell

Right Supplementary Motor Area
Example Cell

Figure VI-6: Pre-SMA cells responding to physical and perceptual switches*. Two example cells recorded in one patient from pre-SMA in the left hemisphere (top) and right hemisphere (bottom) respectively. Responses are aligned to saccades to a fixation spot when the percept was switching (left column) or when a saccade was made without the percept switching, i.e., when the patient kept following the same fixation spot (right column). The three rows for each cell correspond to three stages of the experiment, where the patient was doing the fixation task while physically switching monocular stimuli were presented (first row), where the patient was doing the fixation task while perceiving switches during binocular rivalry and actively reporting the percept by button press (second row), or where the patient was doing the fixation task while perceiving switches during binocular rivalry without report (third row). Blue and red lines indicate responses when the perceived stimulus was the left eye stimulus or right eye stimulus, respectively (as inferred from eye movement).*

Chapter VII

FUTURE DIRECTIONS

Are conscious switches feedforward or feedback?

We have shown that cells in face patches are modulated by conscious percept during binocular rivalry, but strictly speaking, that does not prove that face patches represent the conscious percept of a face in general. After all, it is possible that activity in the face patches could be explained as passive filter from V1 input, if one postulates that V1 is fully responsible for switches during binocular rivalry. In other words, since in binocular rivalry, the conscious percept appears to switch between the two eyes (despite studies claiming otherwise (KOVAcs et al., 1996; Logothetis et al., 1996)), it may be completely resolved and explained by competition of monocular channels within V1 or LGN (Blake et al., 1980; Blake, 1989). Thus, it would be not much more than a mechanical, intra-cortical shutting of one eye, alternating every few seconds. In this scenario, with face patches simply reflecting the shutting and gating of low-level eye information in V1, it would be far-fetched to interpret our results in the framework of face patches representing consciousness. However, models of competition between monocular populations seem to be at odds with electrophysiological experiments that reported no or only weak traces of modulation by conscious percept in LGN (Lehky and Maunsell, 1996) and V1 (Leopold and Logothetis, 1996; Keliris et al., 2010), respectively, and that found no difference between monocular and binocular neurons. Alternatively, binocular rivalry may be a special case of a more general phenomenon of how the brain constructs a conscious percept. We put forward the hypothesis that this generation of a conscious percept may be tightly linked to the process by which the brain resolves ambiguous input and generates an interpretation of the world that makes sense and can recreate sensory input. Some theories of consciousness in the literature state that the initial feedforward processing is unconscious and automatic, and the signal is required to reach a higher-level region to rise into awareness (Dehaene et al., 2003; Baars, 2005; Lau and Rosenthal, 2011), with some of the theories claiming that the signal additionally needs to be fed back from the higher-level regions to lower-level regions (Hochstein and Ahissar, 2002; Lamme, 2006). Evidence for feedback is indirectly suggested by what can be described as one of the biggest discrepancies between fMRI and electrophysiology literature: in fMRI experiments, V1 strongly reflects the conscious percept during rivalry (Polonsky et al., 2000b; Tong and Engel, 2001). However, electrophysiology studies show that actually only a small proportion of V1 cells (~18%) is modulated by conscious percept at all (Leopold and Logothetis, 1996), and only higher-level regions encode conscious percept reliably (~90% of cells in IT) (Sheinberg and Logothetis, 1997). For the extracellular local field potential (LFP), it turned out that low-frequency power followed conscious perception whereas higher frequency power does not (Gail et al., 2004; Wilke et al., 2006; Maier et al., 2008). This discrepancy can be resolved with the knowledge that high-frequency LFP power reflects more the spiking of neurons, whereas low-frequency power and fMRI signals are more affected by input from neuromodulatory systems. Thus, the observed modulation of fMRI signal may reflect feedback from higher-order cortices or subcortical targets (Belitski et al., 2008; Magri et al., 2012).

To distinguish these different scenarios, we need to find out where and how switches in binocular rivalry are generated. Unfortunately, virtually all studies to date have merely investigated which neurons or brain regions correlate with a given conscious percept, but this has given us little insight into the mechanisms of how a conscious percept is constructed across the brain. Having shown that face patches encode the specific conscious percept of a face puts us in a unique position to go beyond studying correlates and dissect the neural mechanisms of how a new conscious percept is constructed and dynamically propagated across the brain. We are in a unique position because the face patch system is comprised of nodes in a functional hierarchy, that are directly anatomically connected and all encode the same category of object. This is critical because it allows us to study the dynamics of information flow as a new conscious percept is established and propagated through the hierarchy. E.g., if new conscious percepts are generated through a top-down mechanism, we would be able to tell, as in that case the higher-level nodes of the hierarchy should show signatures of the switch before the lower-level nodes, and vice-versa for a bottom-up mechanism. By simultaneously recording from nodes at two different levels of the hierarchy, we are thus able to distinguish between possible mechanisms of conscious percept generation. Moreover, we are now equipped with the right tools to study it. While traditionally we have been a single tungsten electrode lab, I had a hunch that for my purpose we would need to record from large populations of neurons, and I hence contributed a handful of high-yield recording techniques to the lab by learning the surgery and successfully implanting new types of electrodes such as V-probes, S-probes, Utah arrays, brush arrays, and Neuropixels electrodes, thereby setting up systems with the capacity to record from up to 512 channels simultaneously.

In Fig. VI-4, we showed an example of the world's first recording with 128-ch. Neuropixels protoype probes designed specifically for primates. We have since started to record from multiple face patches simultaneously, inserting a Neuropixels probe into face patch ML and an S-probe (for technical reasons regarding the length of the probe) into face patch AM. Through this experiment, we are hoping to dissect the mechanisms of binocular rivalry and determine whether perceptual switches are propagated through a feedforward or feedback mechanism. Preliminary data is shown in Fig. VII-1, but please note that this is still early stages of the experiment, and results have yet to be confirmed or disconfirmed. The rationale is that if conscious percepts are generated and propagated through a feedforward mechanism, we should see activity in ML, the lower node of the hierarchy first, followed by activity in AM, which receives input from ML. In case of a feedback mechanism, we should see the reverse.

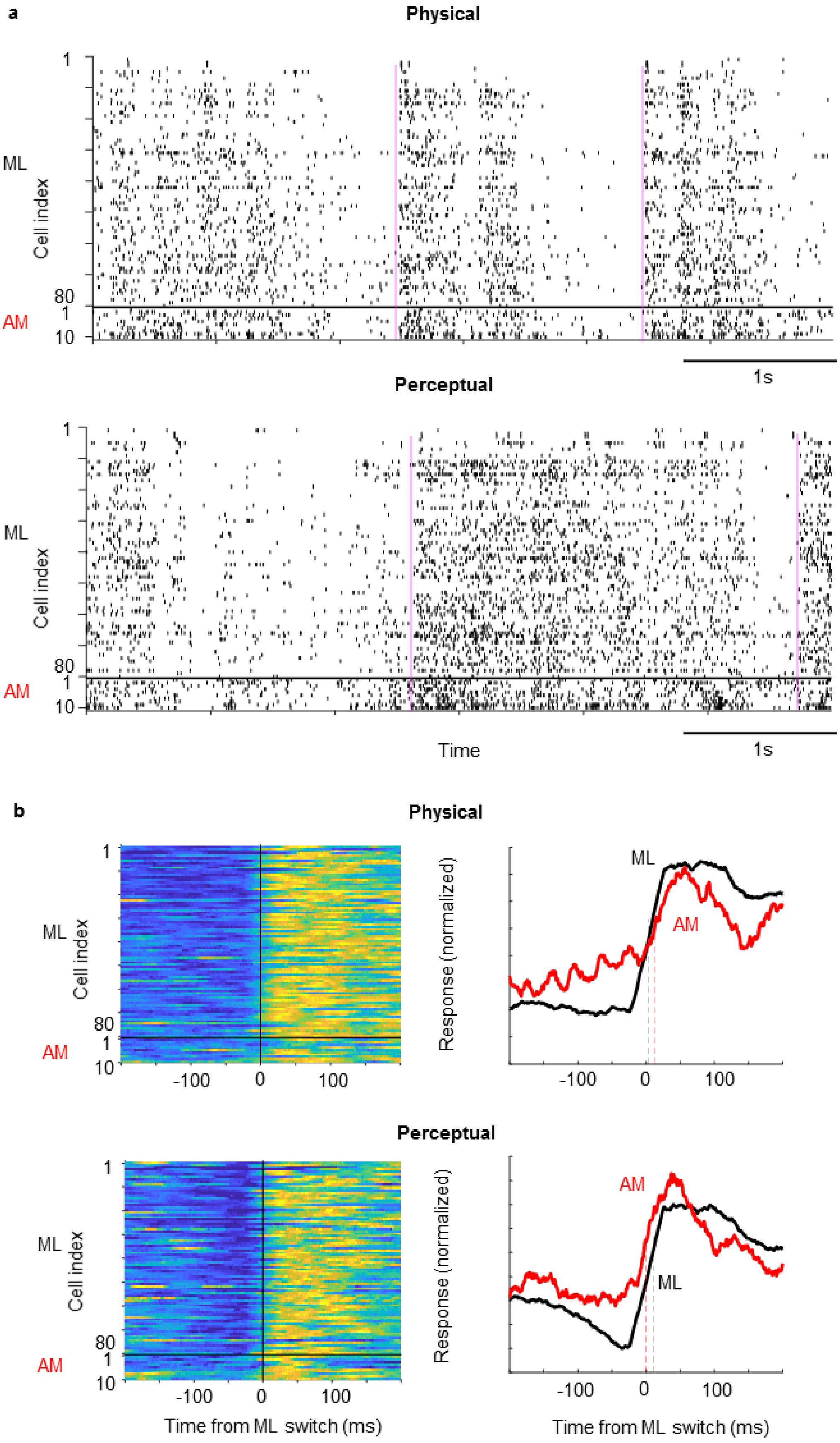

a
Physical
1
ML
Cell index
80
AM
1
10
1s
Perceptual
1
ML
Cell index
80
AM
1
10
Time
1s
b
Physical
1
ML
Cell index
80
AM
1
10
-100
0
100
Response (normalized)
ML
AM
-100
0
100
Perceptual
1
ML
Cell index
80
AM
1
10
-100
0
100
Time from ML switch (ms)
Response (normalized)
AM
ML
-100
0
100
Time from ML switch (ms)

***Figure VII-1: Latencies of switches in ML vs AM**. (**a**) Example population time course of face cells recorded simultaneously with Neuropixels probe in ML and S-probe in AM for the physical condition (top) and perceptual condition (bottom). Cells on top of black horizontal line are from ML, cells below are from AM. Magenta vertical lines indicate switches of ML activity from silent to active, i.e., putative switches from object to face. (**b**) Average activity in ML and AM aligned to switches inferred by ML activity for physical condition (top) and perceptual condition (bottom). Left column shows activity averaged across ML switches for each cell, right column also averages across cells. Dashed vertical lines in the right plot indicate latencies at which activity in ML (black) or AM (red) reached half of the peak value.*

Fig. VII-1a shows example population time course snippets from ML and AM recorded simultaneously. We inferred switches from ML activity alone, by finding time points where face cells changed from being silent to being active (magenta). These are putative time points where the stimulus or percept possibly switched from object to face. We then looked at AM activity around these ML-inferred switches. In the example, it looks as if for physical switches, AM cells fired slightly later than ML cells, whereas for the perceptual switch, AM cells started firing a slightly earlier. We averaged activity across all ML-inferred switches of a session to compare response latency of ML and AM (Fig. VII-2b). For physical switches, ML cells average activity reached its half-peak before AM, as is expected since AM is later in the hierarchy of visual processing and receives direct feedforward input from ML Intriguingly, for perceptual switches, the converse was true: AM reached its half-peak earlier than ML. Thus, this preliminary data is consistent with a top-down mechanism for switches of conscious percept. Further experiments are necessary to see if this result can be reproduced, trace down whether there is a source of conscious percept generation in the brain, and determine the causal role of different nodes of the network.

Future experiments on mechanisms of consciousness

The experiment of the previous section, recording from ML and AM simultaneously during binocular rivalry, holds the potential of yielding one important puzzle piece of the mechanisms of consciousness: whether switches in conscious percept during rivalry are generated through a bottom-up or top-down mechanism. E.g., if we find that AM does indeed show activity before ML during perceptual switches, that would suggest that switches originate from a region higher than ML. Possible candidates would include e.g., the prefrontal cortex or hippocampus. Where do we go from there, and how can we find this source (assuming there is one region that has absolute authority in conscious percept selection, rather than the whole brain acting as a recurrent network to converge to a solution)? We can use the same strategy that we used for recording from ML and AM to trace down the origin of the signal, by recording from ML/AM to detect switches and recording from other brain regions, e.g., prefrontal cortex or hippocampus at the same time and comparing latencies. This assumes that we can find a stimulus that drives both ML/AM and the new region, but there are face patches outside of inferotemporal cortex, including in perirhinal, prefrontal, and more (see Chapter II, section Anatomical organization) that can be targeted. Assume we find that a given region X has a latency lower than ML and AM. Then we can look at the regions that are connected to X, by electrically stimulating region X in an fMRI experiment, and see which other regions light up, and see if one of those regions has even lower latency. By repeating this process, we should eventually arrive at the source of conscious percept construction, if there is one.

Performing simultaneous recordings can show us where a new conscious percept first becomes explicit, but that does not prove a causal role of the region. Once the putative source of conscious percept generation is identified, it will be fascinating to find out what happens if we inactivate it or stimulate it. We hypothesize that after inactivation, the animal will fail to show switches of conscious percept or re-interpretations of the world, both neurally and behaviorally. There may also be other surprising side effects accompanying inactivation of the site, besides lack of switches in the binocular rivalry paradigm, if we go as far as conjecturing that we create a philosophical zombie monkey.

Another interesting possible experiment involves stimulation and recording at the same time to answer the question of whether a given region can causally bias the percept and behavior during binocular rivalry: the idea is to stimulate a higher-level association region, such as face patch PR in perirhinal cortex or in hippocampus to evoke the memory of a face, while recording from face patches in inferotemporal cortex using an ultra-high yield probe such as Neuropixels. Assuming the evoked memory causes the perceptual visualization of a face, and this visualization is represented in face patches in inferotemporal cortex, we should be able to reconstruct the face the monkey was seeing during stimulation from the recorded face cell activity, since we have previously cracked the code of face identity in IT face patches and can reconstruct the face a monkey was seeing from a hundred face cells. Inserting the stimulation electrode into different sub-regions may evoke percepts of different faces that we should be able to reconstruct on a single-trial basis by recording from hundreds of neurons with Neuropixels probes. Assuming we can reconstruct what face is perceived due to stimulation, we can then make binocular rivalry stimuli out of exactly that face in one eye, and another face or object in another eye. Using the no-report paradigm to infer his percept, we can then find out whether stimulation of the higher-level site will bias the animal to perceive the face evoked by stimulation, i.e., whether he will follow the fixation spot in the eye of that face more often. If so, this would imply that this region has a causal influence on conscious percept generation and suggest the possibility that this influence feeds back all the way to V1 to select the stimulus from the respective eye, since it also enhanced the fixation spot of that eye. Inactivating other regions while stimulating may further reveal which regions play a causally necessary role in propagating the conscious percept.

Summary of thesis main results

- In bistable illusions, switches of conscious percept appear to be propagated across multiple layers of representation to generate a consistent interpretation.

- The macaque face patch system may provide a key for unlocking the mystery of the neural mechanisms of conscious perception.

- Consistent border-ownership cells are hard to find in random locations of retinotopic cortex, but the average population signal encodes segmentation for artificial and natural stimuli. For stimuli with subjective contours, border-ownership signal latencies are longer.

- Using fMRI, one can identify segmentation-related hot spots. When recording in these cells, one can find clusters of cells that do signal segmentation consistently across stimuli. The segmentation of an image can be decoded from populations of neurons in segmentation hot spots.

- One can make face patches respond to Mooney faces by adding an outline or animating them.

- When trying to distill representations of conscious percept, one needs to distill conscious percept from several confounding factors.

- Conscious percept is encoded in inferotemporal face patches and can be decoded from a population of neurons.

- Recording or perturbing several nodes of the cortical processing hierarchy simultaneously has the potential to dissect the mechanisms of conscious percept generation and propagation, e.g., whether the mechanism is top-down or bottom-up

BIBLIOGRAPHY

Adachi I, Chou DP, Hampton RR (2009) Thatcher effect in monkeys demonstrates conservation of face perception across primates. Current Biology 19:1270-1273.

Adolphs R, Tranel D, Damasio H, Damasio A (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature 372:669.

Afraz A, Boyden ES, DiCarlo JJ (2015) Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. Proceedings of the National Academy of Sciences 112:6730-6735.

Afraz S-R, Kiani R, Esteky H (2006) Microstimulation of inferotemporal cortex influences face categorization. Nature 442:692.

Aparicio PL, Issa EB, DiCarlo JJ (2016) Neurophysiological organization of the middle face patch in macaque inferior temporal cortex. Journal of Neuroscience 36:12729-12745.

Arcaro MJ, Livingstone MS (2017) A hierarchical, retinotopic proto-organization of the primate visual system at birth. Elife 6.

Arrington CM, Carr TH, Mayer AR, Rao SM (2000) Neural mechanisms of visual attention: object-based selection of a region in space. Journal of cognitive neuroscience 12:106-117.

Aru J, Bachmann T, Singer W, Melloni L (2012) Distilling the neural correlates of consciousness. Neuroscience & Biobehavioral Reviews 36:737-746.

Baars BJ (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Progress in brain research 150:45-53.

Baek K, Sajda P (2005) Inferring figure-ground using a recurrent integrate-and-fire neural circuit. Neural Systems and Rehabilitation Engineering, IEEE Transactions on 13:125-130.

Bahrick HP, Bahrick PO, Wittlinger RP (1975) Fifty years of memory for names and faces: A cross-sectional approach. Journal of experimental psychology: General 104:54.

Bakin JS, Nakayama K, Gilbert CD (2000) Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. The Journal of Neuroscience 20:8188-8198.

Bao P, Tsao DY (2018) Representation of multiple objects in macaque category-selective areas. Nature communications 9:1774.

Bao P, She L, Mcgill M, Tsao DY (2019) A map of object space in primate inferotemporal cortex. In: Society for Neuroscience. Chicago, IL.

Barat E, Wirth S, Duhamel J-R (2018) Face cells in orbitofrontal cortex represent social categories. Proceedings of the National Academy of Sciences 115:E11158-E11167.

Baylis G, Rolls ET, Leonard C (1985) Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. Brain research 342:91-102.

Baylis GC, Rolls ET, Leonard C (1987) Functional subdivisions of the temporal lobe neocortex. Journal of Neuroscience 7:330-342.

Belitski A, Gretton A, Magri C, Murayama Y, Montemurro MA, Logothetis NK, Panzeri S (2008) Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. Journal of Neuroscience 28:5696-5709.

Blake R (1989) A neural theory of binocular rivalry. Psychological review 96:145.

Blake R, Westendorf DH, Overton R (1980) What is suppressed during binocular rivalry? Perception 9:223-231.

Blake R, Brascamp J, Heeger DJ (2014) Can binocular rivalry reveal neural correlates of consciousness? Philosophical Transactions of the Royal Society of London B: Biological Sciences 369:20130211.

Block N (1995) On a confusion about a function of consciousness. Behavioral and brain sciences 18:227-247.

Block N (2019) What is wrong with the no-report paradigm and how to fix it. Trends in cognitive sciences.

Block N (2020) Finessing the bored monkey problem.

Boly M, Massimini M, Tsuchiya N, Postle BR, Koch C, Tononi G (2017) Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. Journal of Neuroscience 37:9603-9613.

Borra E, Ichinohe N, Sato T, Tanifuji M, Rockland KS (2010) Cortical connections to area TE in monkey: hybrid modular and distributed organization. Cereb Cortex 20:257-270.

Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45:602-607.

Breitmeyer BG, Hoar WS, Randall D, Conte FP (1984) Visual masking: An integrative approach: Clarendon Press.

Bruce C (1982) Face recognition by monkeys: absence of an inversion effect. Neuropsychologia 20:515-521.

Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. Journal of neurophysiology 46:369-384.

Bruce V, Young A (1986) Understanding face recognition. British journal of psychology 77:305-327.

Calder AJ, Nummenmaa L (2007) Face cells: separate processing of expression and gaze in the amygdala. Current biology 17:R371-R372.

Cerf M, Frady EP, Koch C (2009) Faces and text attract gaze independent of the task: Experimental data and computer model. Journal of vision 9:10-10.

Chang L, Tsao DY (2017) The code for facial identity in the primate brain. Cell 169:1013-1028. e1014.

Chang L, Bao P, Tsao DY (2017) The representation of colored objects in macaque color patches. Nat Commun 8:2064.

Chen M, Yan Y, Gong X, Gilbert CD, Liang H, Li W (2014) Incremental integration of global contours through interplay between visual cortical areas. Neuron 82:682-694.

Chessick R, D (2008) The Blackwell Companion to Consciousness, edited by Max Velmans and Susan Schneider, Malden, MA, Blackwell Publishing Ltd., 2007. Journal of the American Academy of Psychoanalysis and Dynamic Psychiatry 36:769-773.

Clark V, Keil K, Maisog JM, Courtney S, Ungerleider LG, Haxby JV (1996) Functional magnetic resonance imaging of human visual cortex during face matching: a comparison with positron emission tomography. Neuroimage 4:1-15.

Cootes TF, Edwards GJ, Taylor CJ (1998) Active appearance models. In: European conference on computer vision, pp 484-498: Springer.

Coulon M, Deputte BL, Heyman Y, Baudoin C (2009) Individual recognition in domestic cattle (Bos taurus): evidence from 2D-images of heads from different breeds. PLoS One 4:e4441.

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386:608.

Craft E, Schütze H, Niebur E, Von Der Heydt R (2007) A neural model of figure–ground organization. Journal of neurophysiology 97:4310-4326.

Crick FC, Koch C (2005) What is the function of the claustrum? Philosophical Transactions of the Royal Society B: Biological Sciences 360:1271-1279.

Cuaya LV, Hernández-Pérez R, Concha L (2016) Our faces in the dog's brain: Functional imaging reveals temporal cortex activation during perception of human faces. PloS one 11:e0149431.

de Beeck HPO, Baker CI, DiCarlo JJ, Kanwisher NG (2006) Discrimination training alters object representations in human extrastriate cortex. Journal of Neuroscience 26:13025-13036.

Deen B, Richardson H, Dilks DD, Takahashi A, Keil B, Wald LL, Kanwisher N, Saxe R (2017) Organization of high-level visual cortex in human infants. Nature communications 8:13995.

Dehaene S, Sergent C, Changeux J-P (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proceedings of the National Academy of Sciences 100:8520-8525.

Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, Dehaene-Lambertz G, Kolinsky R, Morais J, Cohen L (2010) How learning to read changes the cortical networks for vision and language. science 330:1359-1364.

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. Journal of Neuroscience 4:2051-2062.

Desimone R, Wessinger M, Thomas L, Schneider W (1990) Attentional control of visual perception: cortical and subcortical mechanisms. In: Cold Spring Harbor symposia on quantitative biology, pp 963-971: Cold Spring Harbor Laboratory Press.

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends in cognitive sciences 11:333-341.

Doi H, Koga T, Shinohara K (2009) 18-Month-olds can perceive Mooney faces. Neuroscience Research 64:317-322.

Driver J, Baylis GC (1996) Edge-assignment and figure–ground segmentation in short-term visual matching. Cognitive psychology 31:248-306.

Dubois J, de Berker AO, Tsao DY (2015) Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. Journal of Neuroscience 35:2791-2802.

Dutta B, Andrei A, Harris T, Lopez C, O’Callahan J, Putzeys J, Raducanu B, Severi S, Stavisky S, Trautmann E (2019) The Neuropixels probe: A CMOS based integrated microsystems platform for neuroscience and brain-computer interfaces.

In: 2019 IEEE International Electron Devices Meeting (IEDM), pp 10.11. 11-10.11. 14: IEEE.

Edelstein L, Denaro F (2004) The claustrum: a historical review of its anatomy, physiology, cytochemistry and functional significance. Pathology 104:368,415,434,556,557.

Erickson CA, Jagadeesh B, Desimone R (2000) Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. Nature neuroscience 3:1143.

Ettlinger G, Wilson W (1990) Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms. Behavioural brain research 40:169-192.

Fahy F, Riches I, Brown M (1993) Neuronal activity related to visual recognition memory: long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. Experimental Brain Research 96:457-472.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.

Fernández-Miranda JC, Rhoton Jr AL, Kakizawa Y, Choi C, Álvarez-Linera J (2008) The claustrum and its projection system in the human brain: a microsurgical and tractographic anatomical study.

Finkel LH, Sajda P (1992) Object discrimination based on depth-from-occlusion. Neural Computation 4:901-921.

Fisch L, Privman E, Ramot M, Harel M, Nir Y, Kipervasser S, Andelman F, Neufeld MY, Kramer U, Fried I (2009) Neural “ignition”: enhanced activation linked to perceptual awareness in human ventral stream visual cortex. Neuron 64:562-574.

Fisher C, Freiwald WA (2015) Contrasting specializations for facial motion within the macaque face-processing system. Current Biology 25:261-266.

Fowlkes CC, Martin DR, Malik J (2007) Local figure–ground cues are valid for natural images. Journal of Vision 7:2-2.

Frässle S, Sommer J, Jansen A, Naber M, Einhäuser W (2014) Binocular rivalry: frontal activity relates to introspection and action but not to perception. Journal of Neuroscience 34:1738-1747.

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

Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. Nature neuroscience 12:1187.

Fried I, MacDonald KA, Wilson CL (1997) Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. Neuron 18:753-765.

Fries P, Roelfsema PR, Engel AK, König P, Singer W (1997) Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. Proceedings of the National Academy of Sciences 94:12699-12704.

Friston K (2009) The free-energy principle: a rough guide to the brain? Trends in cognitive sciences 13:293-301.

Furl N, Hadj-Bouziane F, Liu N, Averbeck BB, Ungerleider LG (2012) Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. Journal of Neuroscience 32:15952-15962.

Gail A, Brinksmeyer HJ, Eckhorn R (2004) Perception-related modulations of local field potential power and coherence in primary visual cortex of awake monkey during binocular rivalry. Cerebral Cortex 14:300-313.

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects. Nature neuroscience 2:568.

Gelbard-Sagiv H, Mudrik L, Hill MR, Koch C, Fried I (2018) Human single neuron activity precedes emergence of conscious perception. Nature communications 9:2057.

Gilbert CD, Li W (2013) Top-down influences on visual processing. Nature Reviews Neuroscience 14:350-363.

Goren CC, Sarty M, Wu PY (1975) Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics 56:544-549.

Gothard KM, Brooks KN, Peterson MA (2009) Multiple perceptual strategies used by macaque monkeys for face recognition. Anim Cogn 12:155-167.

Grill-Spector K, Kanwisher N (2005) Visual recognition as soon as you know it is there, you know what it is. Psychological Science 16:152-160.

Grimaldi P, Saleem KS, Tsao D (2016) Anatomical connections of the functionally defined “face patches” in the macaque monkey. Neuron 90:1325-1342.

Grosof D, Shapley R, Hawken M (1993) Macaque V1 neurons can signal illusory contours. Nature 365:550-552.

Gross CG (2006) Charles G. Gross. In: History of Neuroscience in Autobiography (Albright T, Squire LR, eds): Oxford University Press.

Gross CG, Rocha-Miranda Cd, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the Macaque. Journal of neurophysiology 35:96-111.

Hadj-Bouziane F, Bell AH, Knusten TA, Ungerleider LG, Tootell RB (2008) Perception of emotional expressions is independent of face selectivity in monkey inferior temporal cortex. Proceedings of the National Academy of Sciences 105:5591-5596.

Hasson U, Harel M, Levy I, Malach R (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. Neuron 37:1027-1041.

Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends in cognitive sciences 4:223-233.

Haxby JV, Ungerleider LG, Clark VP, Schouten JL, Hoffman EA, Martin A (1999) The effect of face inversion on activity in human neural systems for face and object perception. Neuron 22:189-199.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430.

Heitger F, von der Heydt R, Kübler O (1994) A computational model of neural contour processing: Figure-ground segregation and illusory contours. In: From Perception to Action Conference, 1994., Proceedings, pp 181-192: IEEE.

Hesse JK, Tsao DY (2016) Consistency of border-ownership cells across artificial stimuli, natural stimuli, and stimuli with ambiguous contours. Journal of Neuroscience 36:11338-11349.

Hesse JK, Tsao DY (2020) Representation of conscious percept without report in the macaque face patch network. bioRxiv.

Hesse JK, Wadia V, Rutishauser U, Tsao DY (2019) Neural correlates of perceptual switches in binocular rivalry without active report. In: Society for Neuroscience. Chicago, IL.

Heywood C, Cowey A (1992) The role of the ‘face-cell’area in the discrimination and recognition of faces by monkeys. Phil Trans R Soc Lond B 335:31-38.

Hicks RD (2015) Aristotle De Anima: Cambridge University Press.

Hochstein S, Ahissar M (2002) View from the top: Hierarchies and reverse hierarchies in the visual system. Neuron 36:791-804.

Hoeks B, Levelt WJ (1993) Pupillary dilation as a measure of attention: A quantitative system analysis. Behavior Research Methods, Instruments, & Computers 25:16-26.

Hohwy J, Roepstorff A, Friston K (2008) Predictive coding explains binocular rivalry: An epistemological review. Cognition 108:687-701.

Huang Y, Rao RP (2011) Predictive coding. Wiley Interdisciplinary Reviews: Cognitive Science 2:580-593.

Hung C-C, Yen CC, Ciuchta JL, Papoti D, Bock NA, Leopold DA, Silva AC (2015) Functional mapping of face-selective regions in the extrastriate visual cortex of the marmoset. Journal of Neuroscience 35:1160-1172.

Inagaki M, Fujita I (2011) Reference frames for spatial frequency in face representation differ in the temporal visual cortex and amygdala. Journal of Neuroscience 31:10371-10379.

Issa EB, DiCarlo JJ (2012) Precedence of the eye region in neural processing of faces. Journal of Neuroscience 32:16666-16682.

Janssens T, Zhu Q, Popivanov ID, Vanduffel W (2014) Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. Journal of Neuroscience 34:10156-10167.

Jehee JF, Lamme VA, Roelfsema PR (2007) Boundary assignment in a recurrent network architecture. Vision research 47:1153-1165.

Johnson MH, Dziurawiec S, Ellis H, Morton J (1991) Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition 40:1-19.

Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydın Ç (2017) Fully integrated silicon probes for high-density recording of neural activity. Nature 551:232.

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

Kapoor V, Besserve M, Logothetis NK, Panagiotaropoulos TI (2018) Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. Communications biology 1:215.

Kapoor V, Dwarakanath A, Safavi S, Werner J, Besserve M, Panagiotaropoulos TI, Logothetis NK (2020) Decoding the contents of consciousness from prefrontal ensembles. bioRxiv.

Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. Nature neuroscience 22:974.

Keliris GA, Logothetis NK, Tolias AS (2010) The role of the primary visual cortex in perceptual suppression of salient visual stimuli. Journal of Neuroscience 30:12353-12365.

Kendrick K, Baldwin B (1987) Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces. Science 236:448-450.

Khuvis S, Yeagle E, Mehta A (2017) Diverse response properties of face-selective cells in the human fusiform face area. In: Society for Neuroscience Conference, p 192.106. Washington DC.

Khuvis S, Yeagle EM, Norman Y, Grossman S, Malach R, Mehta AD (2018) Face-selective units in human ventral temporal cortex reactivate during free recall. BioRxiv:487686.

Kiani R, Esteky H, Tanaka K (2005) Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. Journal of neurophysiology 94:1587-1596.

Koch C, Tsuchiya N (2007) Attention and consciousness: two distinct brain processes. Trends in cognitive sciences 11:16-22.

Koch C, Massimini M, Boly M, Tononi G (2016) Neural correlates of consciousness: progress and problems. Nature Reviews Neuroscience 17:307.

Kogo N, Wagemans J (2013) The “side” matters: How configurality is reflected in completion. Cognitive neuroscience 4:31-45.

Kogo N, Strecha C, Van Gool L, Wagemans J (2010) Surface construction by a 2-D differentiation–integration process: A neurocomputational model for perceived border ownership, depth, and lightness in Kanizsa figures. Psychological review 117:406.

Kolster H, Janssens T, Orban GA, Vanduffel W (2014) The retinotopic organization of macaque occipitotemporal cortex anterior to V4 and caudoventral to the middle temporal (MT) cluster. Journal of Neuroscience 34:10168-10191.

Konorski J (1967) Integrative activity of the brain.

Kornblith S, Tsao DY (2017) How thoughts arise from sights: inferotemporal and prefrontal contributions to vision. Current opinion in neurobiology 46:208-218.

Kornblith S, Cheng X, Ohayon S, Tsao DY (2013) A network for scene processing in the macaque temporal lobe. Neuron 79:766-781.

KovAcs G, Vogels R, Orban GA (1995) Cortical correlate of pattern backward masking. Proceedings of the National Academy of Sciences 92:5587-5591.

KOVAcs I, Papathomas TV, Yang M, Fehér Á (1996) When the brain changes its mind: Interocular grouping during binocular rivalry. Proceedings of the National Academy of Sciences 93:15508-15511.

Ku S-P, Tolias AS, Logothetis NK, Goense J (2011) fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. Neuron 70:352-362.

Kumar S, Popivanov ID, Vogels R (2017) Transformation of Visual Representations Across Ventral Stream Body-selective Patches. Cereb Cortex:1-15.

Kuraoka K, Nakamura K (2007) Responses of single neurons in monkey amygdala to facial and vocal emotions. Journal of Neurophysiology 97:1379-1387.

Lafer-Sousa R, Conway BR (2013) Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat Neurosci 16:1870-1878.

Lamme VA (2006) Towards a true neural stance on consciousness. Trends in cognitive sciences 10:494-501.

Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends in neurosciences 23:571-579.

Landi SM, Freiwald WA (2017) Two areas for familiar face recognition in the primate brain. Science 357:591-595.

Lau H, Rosenthal D (2011) Empirical support for higher-order theories of conscious awareness. Trends in cognitive sciences 15:365-373.

Lee TS, Nguyen M (2001) Dynamics of subjective contour formation in the early visual cortex. Proceedings of the National Academy of Sciences 98:1907-1911.

Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. JOSA A 20:1434-1448.

Lehky SR, Maunsell JH (1996) No binocular rivalry in the LGN of alert macaque monkeys. Vision research 36:1225-1234.

Leibo JZ, Mutch J, Poggio T (2011) Why the brain separates face recognition from object recognition. In: Advances in neural information processing systems, pp 711-719.

Leibo JZ, Liao Q, Anselmi F, Freiwald WA, Poggio T (2017) View-tolerant face recognition and Hebbian learning imply mirror-symmetric neural tuning to head orientation. Current Biology 27:62-67.

Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. Nature 379:549.

Leopold DA, Plettenberg HK, Logothetis NK (2002) Visual processing in the ketamine-anesthetized monkey. Experimental brain research 143:359-372.

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

Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center–periphery organization of human object areas. Nature neuroscience 4:533.

Logothetis NK, Schall JD (1989) Neuronal correlates of subjective visual perception. Science 245:761-763.

Logothetis NK, Leopold DA, Sheinberg DL (1996) What is rivalling during binocular rivalry? Nature 380:621-624.

Magri C, Schridde U, Murayama Y, Panzeri S, Logothetis NK (2012) The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. Journal of Neuroscience 32:1395-1407.

Maier A, Wilke M, Aura C, Zhu C, Frank QY, Leopold DA (2008) Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. Nature neuroscience 11:1193.

Malach R, Reppas J, Benson R, Kwong K, Jiang H, Kennedy W, Ledden P, Brady T, Rosen B, Tootell R (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proceedings of the National Academy of Sciences 92:8135-8139.

Marciniak K, Atabaki A, Dicke PW, Thier P (2014) Disparate substrates for head gaze following and face perception in the monkey superior temporal sulcus. Elife 3.

Marr D (1982) Vision: A computational investigation into the human representation and processing of visual information. New York, NY, USA: Henry Holt and Co. Inc June.

Maunsell JH, Gibson JR (1992) Visual response latencies in striate cortex of the macaque monkey. Journal of Neurophysiology 68:1332-1344.

Maunsell JH, Treue S (2006) Feature-based attention in visual cortex. Trends in neurosciences 29:317-322.

McDermott J (2004) Psychophysics with junctions in real images. Perception 33:1101-1127.

McMahon DB, Jones AP, Bondar IV, Leopold DA (2014) Face-selective neurons maintain consistent visual responses across months. Proceedings of the National Academy of Sciences 111:8251-8256.

McMahon DB, Russ BE, Elnaiem HD, Kurnikova AI, Leopold DA (2015) Single-unit activity during natural vision: diversity, consistency, and spatial sensitivity among AF face patch neurons. Journal of Neuroscience 35:5537-5548.

Medawar PB (2008) Advice to a young scientist: Basic Books.

Meng M, Tong F (2004) Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. Journal of vision 4:2-2.

Minxha J, Mamelak AN, Rutishauser U (2018) Surgical and electrophysiological techniques for single-neuron recordings in human epilepsy patients. In: Extracellular recording approaches, pp 267-293: Springer.

Minxha J, Mosher C, Morrow JK, Mamelak AN, Adolphs R, Gothard KM, Rutishauser U (2017) Fixations gate species-specific responses to free viewing of faces in the human and macaque amygdala. Cell reports 18:878-891.

Mitchell JF, Stoner GR, Reynolds JH (2004) Object-based attention determines dominance in binocular rivalry. Nature 429:410.

Moeller S, Freiwald WA, Tsao DY (2008) Patches with links: a unified system for processing faces in the macaque temporal lobe. Science 320:1355-1359.

Moeller S, Crapse T, Chang L, Tsao DY (2017) The effect of face patch microstimulation on perception of faces and objects. Nature neuroscience 20:743-752.

Mosher CP, Zimmerman PE, Gothard KM (2014) Neurons in the monkey amygdala detect eye contact during naturalistic social interactions. Current Biology 24:2459-2464.

Naber M, Frässle S, Einhäuser W (2011) Perceptual rivalry: reflexes reveal the gradual nature of visual awareness. PLoS One 6:e20910.

Nakamura K, Mikami A, Kubota K (1992) Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. Journal of Neurophysiology 67:1447-1463.

Nakayama K, He ZJ, Shimojo S (1995) Visual surface representation: A critical link between lower-level and higher-level vision. Visual cognition: An invitation to cognitive science 2:1-70.

Newport C, Wallis G, Reshitnyk Y, Siebeck UE (2016) Discrimination of human faces by archerfish (Toxotes chatareus). Scientific reports 6:27523.

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

Ohayon S, Tsao DY (2012) MR-guided stereotactic navigation. J Neurosci Methods 204:389-397.

Ohayon S, Freiwald WA, Tsao DY (2012) What makes a cell face selective? The importance of contrast. Neuron 74:567-581.

Olshausen BA, Anderson CH, Van Essen DC (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. Journal of Neuroscience 13:4700-4719.

Ooi TL, He ZJ (1999) Binocular rivalry and visual awareness: The role of attention. Perception 28:551-574.

Overgaard M, Fazekas P (2016) Can no-report paradigms extract true correlates of consciousness? Trends in cognitive sciences 20:241-242.

Overman Jr WH, Doty RW (1982) Hemispheric specialization displayed by man but not macaques for analysis of faces. Neuropsychologia 20:113-128.

Pachitariu M, Steinmetz NA, Kadir SN, Carandini M, Harris KD (2016) Fast and accurate spike sorting of high-channel count probes with KiloSort. In: Advances in neural information processing systems, pp 4448-4456.

Pack CC, Born RT, Livingstone MS (2003) Two-dimensional substructure of stereo and motion interactions in macaque visual cortex. Neuron 37:525-535.

Panagiotaropoulos TI, Deco G, Kapoor V, Logothetis NK (2012) Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. Neuron 74:924-935.

Park SH, Russ BE, McMahon DB, Koyano KW, Berman RA, Leopold DA (2017) Functional subpopulations of neurons in a macaque face patch revealed by single-unit fMRI mapping. Neuron 95:971-981. e975.

Parker MP (1999) The archaeology of death and burial: Sutton.

Parr L, Winslow J, Hopkins W (1999) Is the inversion effect in rhesus monkeys face-specific? Animal Cognition 2:123-129.

Parr LA, Heintz M, Pradhan G (2008) Rhesus monkeys (Macaca mulatta) lack expertise in face processing. J Comp Psychol 122:390-402.

Parvizi J, Jacques C, Foster BL, Witthoft N, Rangarajan V, Weiner KS, Grill-Spector K (2012) Electrical stimulation of human fusiform face-selective regions distorts face perception. J Neurosci 32:14915-14920.

Pascalis O, Bachevalier J (1998) Face recognition in primates: a cross-species study. Behavioural processes 43:87-96.

Pearson R, Brodal P, Gatter K, Powell T (1982) The organization of the connections between the cortex and the claustrum in the monkey. Brain research 234:435-441.

Petersen SE, Robinson DL, Morris JD (1987) Contributions of the pulvinar to visual spatial attention. Neuropsychologia 25:97-105.

Peterson MA (1999) What's in a stage name? Comment on Vecera and O'Reilly (1998).

Peterson MA, Gibson BS (1993) Shape recognition inputs to figure-ground organization in three-dimensional displays. Cognitive Psychology 25:383-429.

Peterson MA, Gibson BS (1994) Must figure-ground organization precede object recognition? An assumption in peril. Psychological Science 5:253-259.

Peterson MA, Kim JH (2001) On what is bound in figures and grounds. Visual Cognition 8:329-348.

Pinsk MA, DeSimone K, Moore T, Gross CG, Kastner S (2005) Representations of faces and body parts in macaque temporal cortex: a functional MRI study. Proceedings of the National Academy of Sciences of the United States of America 102:6996-7001.

Pinsk MA, Arcaro M, Weiner KS, Kalkus JF, Inati SJ, Gross CG, Kastner S (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. Journal of neurophysiology 101:2581-2600.

Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N (2011) Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage 56:2356-2363.

Polonsky A, Blake R, Braun J, Heeger DJ (2000a) Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. Nature neuroscience 3:1153.

Polonsky A, Blake R, Braun J, Heeger DJ (2000b) Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. Nature neuroscience 3:1153-1159.

Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2012) Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. Neuroimage 63:723-741.

Premereur E, Taubert J, Janssen P, Vogels R, Vanduffel W (2016) Effective connectivity reveals largely independent parallel networks of face and body patches. Current Biology 26:3269-3279.

Preuschoff K, t Hart BM, Einhauser W (2011) Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. Frontiers in neuroscience 5:115.

Puce A, Allison T, Gore JC, McCarthy G (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. Journal of neurophysiology 74:1192-1199.

Puce A, Allison T, Asgari M, Gore JC, McCarthy G (1996) Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. Journal of neuroscience 16:5205-5215.

Qiu FT, Von Der Heydt R (2005) Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. Neuron 47:155-166.

Qiu FT, Von Der Heydt R (2007) Neural representation of transparent overlay. Nature neuroscience 10:283.

Quiroga RQ, Mukamel R, Isham EA, Malach R, Fried I (2008) Human single-neuron responses at the threshold of conscious recognition. Proceedings of the National Academy of Sciences 105:3599-3604.

Rajimehr R, Young JC, Tootell RB (2009) An anterior temporal face patch in human cortex, predicted by macaque maps. Proceedings of the National Academy of Sciences 106:1995-2000.

Rajimehr R, Bilenko NY, Vanduffel W, Tootell RB (2014) Retinotopy versus face selectivity in macaque visual cortex. Journal of cognitive neuroscience 26:2691-2700.

Ramsden BM, Hung CP, Roe AW (2001) Real and illusory contour processing in area V1 of the primate: a cortical balancing act. Cerebral Cortex 11:648-665.

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience 2:79.

Revonsuo A, Kamppinen M (2013) Consciousness in philosophy and cognitive neuroscience: Psychology Press.

Rey HG, Fried I, Quiroga RQ (2014) Timing of single-neuron and local field potential responses in the human medial temporal lobe. Current Biology 24:299-304.

RielSalvatore J, Clark G, Davidson I, Noble W, DErrico F, Vanhaeren M, Gargett RH, Hovers E, BelferCohen A, Krantz GS (2001) Grave markers: Middle and Early Upper Paleolithic burials and the use of chronotypology in contemporary Paleolithic research. Current Anthropology 42:449-479.

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nature neuroscience 2:1019.

Riesenhuber M, Poggio T (2002) Neural mechanisms of object recognition. Current opinion in neurobiology 12:162-168.

Robinson DL, Petersen SE (1992) The pulvinar and visual salience. Trends in neurosciences 15:127-132.

Romanski LM, Diehl MM (2011) Neurons responsive to face-view in the primate ventrolateral prefrontal cortex. Neuroscience 189:223-235.

Rosenfeld SA, Van Hoesen GW (1979) Face recognition in the rhesus monkey. Neuropsychologia 17:503-509.

Rossion B, Taubert J (2019) What can we learn about human individual face recognition from experimental studies in monkeys? Vision Res 157:142-158.

Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat Neurosci 8:107-113.

Roy A, Shepherd SV, Platt ML (2012) Reversible inactivation of pSTS suppresses social gaze following in the macaque (Macaca mulatta). Social cognitive and affective neuroscience 9:209-217.

Rubin E (1958) Figure and ground. Readings in perception:194-203.

Rubin E (1980) Visuell wahrgenommene figuren: Рипол Классик.

Rutishauser U, Schuman EM, Mamelak AN (2006) Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo. Journal of neuroscience methods 154:204-224.

Rutishauser U, Ross IB, Mamelak AN, Schuman EM (2010) Human memory strength is predicted by theta-frequency phase-locking of single neurons. Nature 464:903-907.

Rutishauser U, Tudusciuc O, Neumann D, Mamelak AN, Heller AC, Ross IB, Philpott L, Sutherling WW, Adolphs R (2011) Single-unit responses selective for whole faces in the human amygdala. Current Biology 21:1654-1660.

Saalmann YB, Kastner S (2011) Cognitive and perceptual functions of the visual thalamus. Neuron 71:209-223.

Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. Science 337:753-756.

Sadagopan S, Zarco W, Freiwald WA (2017) A causal relationship between face-patch activity and face-detection behavior. Elife 6.

Safavi S, Kapoor V, Logothetis NK, Panagiotaropoulos TI (2014) Is the frontal lobe involved in conscious perception? Frontiers in Psychology 5:1063.

Sajda P, Finkel LH (1992) Simulating biological vision with hybrid neural networks. Simulation 59:47-55.

Sakai K, Nishimura H (2006) Surrounding suppression and facilitation in the determination of border ownership. Journal of Cognitive Neuroscience 18:562-579.

Saleem KS, Tanaka K, Rockland KS (1993) Specific and columnar projection from area TEO to TE in the macaque inferotemporal cortex. Cerebral Cortex 3:454-464.

Saleem KS, Miller B, Price JL (2014) Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. The Journal of comparative neurology 522:1641-1690.

Sanghera M, Rolls E, Roper-Hall A (1979) Visual responses of neurons in the dorsolateral amygdala of the alert monkey. Experimental neurology 63:610-626.

Sapountzis P, Schluppeck D, Bowtell R, Peirce JW (2010) A comparison of fMRI adaptation and multivariate pattern classification analysis in visual cortex. Neuroimage 49:1632-1640.

Sato T, Uchida G, Lescroart MD, Kitazono J, Okada M, Tanifuji M (2013) Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. Journal of Neuroscience 33:16642-16656.

Schalk G, Kapeller C, Guger C, Ogawa H, Hiroshima S, Lafer-Sousa R, Saygin ZM, Kamada K, Kanwisher N (2017) Facephenes and rainbows: Causal evidence for functional and anatomical specificity of face and color processing in the human brain. Proc Natl Acad Sci U S A 114:12285-12290.

Schwiedrzik CM, Freiwald WA (2017) High-level prediction signals in a low-level area of the macaque face-processing hierarchy. Neuron 96:89-97. e84.

Sergent J, Ohta S, MACDONALD B (1992) Functional neuroanatomy of face and object processing: a positron emission tomography study. Brain 115:15-36.

She L, Tsao D (2017) Recordings from macaque face and body patches in the upper bank of the superior temporal sulcus reveal strong species selectivity. In: Society for Neuroscience Conference, p 192.104. Washington DC.

She LT, Doris (2018) Face coding in the macaque perirhinal face patch. In: Program No. 307.12. Society for Neuroscience, 2018. .

Sheinberg DL, Logothetis NK (1997) The role of temporal cortical areas in perceptual organization. Proceedings of the National Academy of Sciences 94:3408-3413.

Sherk H (1986) The claustrum and the cerebral cortex. In: Sensory-motor areas and aspects of cortical connectivity, pp 467-499: Springer.

Sherrington CS (1940) Man on his nature.

Sheth BR, Sharma J, Rao SC, Sur M (1996) Orientation maps of subjective contours in visual cortex. Science 274:2110-2115.

Shipp S (2004) The brain circuitry of attention. Trends in cognitive sciences 8:223-230.

Sigala R, Logothetis NK, Rainer G (2011) Own-species bias in the representations of monkey and human face categories in the primate temporal lobe. Journal of neurophysiology 105:2740-2752.

Sinha P (2002) Recognizing complex patterns. nature neuroscience 5:1093.

Srihasam K, Vincent JL, Livingstone MS (2014) Novel domain formation reveals proto-architecture in inferotemporal cortex. Nature neuroscience 17:1776.

Srihasam K, Mandeville JB, Morocz IA, Sullivan KJ, Livingstone MS (2012) Behavioral and anatomical consequences of early versus late symbol training in macaques. Neuron 73:608-619.

Stephan C, Wilkinson A, Huber L (2012) Have we met before? Pigeons recognise familiar human faces. Avian Biology Research 5:75-80.

Sugase-Miyamoto Y, Matsumoto N, Ohyama K, Kawano K (2014) Face inversion decreased information about facial identity and expression in face-responsive neurons in macaque area TE. Journal of Neuroscience 34:12457-12469.

Sugihara T, Qiu FT, von der Heydt R (2011) The speed of context integration in the visual cortex. Journal of neurophysiology 106:374-385.

Sugita Y (2008) Face perception in monkeys reared with no exposure to faces. Proc Natl Acad Sci U S A 105:394-398.

Supèr H, Spekreijse H, Lamme VA (2001) Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nature neuroscience 4:304.

Supèr H, Romeo A, Keil M (2010) Feed-forward segmentation of figure-ground and assignment of border-ownership. PLoS One 5:e10705.

Tan C, Poggio T (2016) Neural tuning size in a model of primate visual processing accounts for three key markers of holistic face processing. PloS one 11:e0150980.

Tanaka JW, Farah MJ (1993) Parts and wholes in face recognition. The Quarterly journal of experimental psychology 46:225-245.

Tanaka K (1996) Inferotemporal cortex and object vision. Annual review of neuroscience 19:109-139.

Tang H, Schrimpf M, Lotter W, Moerman C, Paredes A, Ortega Caro J, Hardesty W, Cox D, Kreiman G (2018a) Recurrent computations for visual pattern completion. Proc Natl Acad Sci U S A 115:8835-8840.

Tang S, Lee TS, Li M, Zhang Y, Xu Y, Liu F, Teo B, Jiang H (2018b) Complex Pattern Selectivity in Macaque Primary Visual Cortex Revealed by Large-Scale Two-Photon Imaging. Curr Biol 28:38-48 e33.

Tanné-Gariépy J, Boussaoud D, Rouiller EM (2002) Projections of the claustrum to the primary motor, premotor, and prefrontal cortices in the macaque monkey. Journal of Comparative Neurology 454:140-157.

Taubert J, Van Belle G, Vanduffel W, Rossion B, Vogels R (2014) The effect of face inversion for neurons inside and outside fMRI-defined face-selective cortical regions. Journal of neurophysiology 113:1644-1655.

Taubert J, Van Belle G, Vanduffel W, Rossion B, Vogels R (2015) Neural correlate of the thatcher face illusion in a monkey face-selective patch. Journal of Neuroscience 35:9872-9878.

Taubert J, Goffaux V, Van Belle G, Vanduffel W, Vogels R (2016) The impact of orientation filtering on face-selective neurons in monkey inferior temporal cortex. Scientific reports 6:21189.

Taubert J, Flessert M, Wardle SG, Basile BM, Murphy AP, Murray EA, Ungerleider LG (2018) Amygdala lesions eliminate viewing preferences for faces in rhesus monkeys. Proceedings of the National Academy of Sciences 115:8043-8048.

Thompson P (1980) Margaret Thatcher: a new illusion. Perception.

Tibbetts EA (2002) Visual signals of individual identity in the wasp Polistes fuscatus. Proceedings of the Royal Society of London B: Biological Sciences 269:1423-1428.

Todorov A, Mandisodza AN, Goren A, Hall CC (2005) Inferences of competence from faces predict election outcomes. Science 308:1623-1626.

Tong F, Engel SA (2001) Interocular rivalry revealed in the human cortical blind-spot representation. Nature 411:195.

Tong F, Meng M, Blake R (2006) Neural bases of binocular rivalry. Trends in cognitive sciences 10:502-511.

Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. Neuron 21:753-759.

Tsao DY, Moeller S, Freiwald WA (2008a) Comparing face patch systems in macaques and humans. Proceedings of the National Academy of Sciences 105:19514-19519.

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

Tsao DY, Schweers N, Moeller S, Freiwald WA (2008b) Patches of face-selective cortex in the macaque frontal lobe. Nature neuroscience 11:877.

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003a) Faces and objects in macaque cerebral cortex. Nature neuroscience 6:989.

Tsao DY, Vanduffel W, Sasaki Y, Fize D, Knutsen TA, Mandeville JB, Wald LL, Dale AM, Rosen BR, Van Essen DC (2003b) Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. Neuron 39:555-568.

Tsuchiya N, Koch C (2005) Continuous flash suppression reduces negative afterimages. Nature neuroscience 8:1096.

Tsuchiya N, Wilke M, Frässle S, Lamme VA (2015) No-report paradigms: extracting the true neural correlates of consciousness. Trends in cognitive sciences 19:757-770.

Tsuchiya N, Frässle S, Wilke M, Lamme V (2016) No-report and report-based paradigms jointly unravel the NCC: response to Overgaard and Fazekas.

Valentine T (1988) Upside-down faces: A review of the effect of inversion upon face recognition. British journal of psychology 79:471-491.

Valenza E, Simion F, Cassia VM, Umiltà C (1996) Face preference at birth. Journal of experimental psychology: Human Perception and Performance 22:892.

Van der Velden J, Zheng Y, Patullo BW, Macmillan DL (2008) Crayfish recognize the faces of fight opponents. PLoS One 3:e1695.

Vecera SP, Farah MJ (1997) Is visual image segmentation a bottom-up or an interactive process? Perception & Psychophysics 59:1280-1296.

Vecera SP, O'reilly RC (1998) Figure-ground organization and object recognition processes: an interactive account. Journal of Experimental Psychology: Human Perception and Performance 24:441.

Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, pp I-I: IEEE.

von der Heydt R (2013) Neurophysiological constraints on models of illusory contours. Cognitive neuroscience 4:49-50.

von der Heydt R, Peterhans E (1989) Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. The Journal of neuroscience 9:1731-1748.

Von der Heydt R, Peterhans E, Baumgartner G (1984) Illusory contours and cortical neuron responses. Science 224:1260-1262.

Wallis JD (2007) Orbitofrontal cortex and its contribution to decision-making. Annu Rev Neurosci 30:31-56.

Wang S, Tudusciuc O, Mamelak AN, Ross IB, Adolphs R, Rutishauser U (2014) Neurons in the human amygdala selective for perceived emotion. Proceedings of the National Academy of Sciences 111:E3110-E3119.

Wilke M, Logothetis NK, Leopold DA (2003) Generalized flash suppression of salient visual targets. Neuron 39:1043-1052.

Wilke M, Logothetis NK, Leopold DA (2006) Local field potential reflects perceptual suppression in monkey visual cortex. Proceedings of the National Academy of Sciences 103:17507-17512.

Wilke M, Mueller K-M, Leopold DA (2009) Neural activity in the visual thalamus reflects perceptual suppression. Proceedings of the National Academy of Sciences 106:9465-9470.

Wilke M, Kagan I, Andersen RA (2013) Effects of pulvinar inactivation on spatial decision-making between equal and asymmetric reward options. Journal of cognitive neuroscience 25:1270-1283.

Wilke M, Turchi J, Smith K, Mishkin M, Leopold DA (2010) Pulvinar inactivation disrupts selection of movement plans. Journal of Neuroscience 30:8650-8659.

Wolfe JM (1984) Reversing ocular dominance and suppression in a single flash. Vision research 24:471-478.

Wright AA, Roberts WA (1996) Monkey and human face perception: Inversion effects for human faces but not for monkey faces or scenes. Journal of Cognitive Neuroscience 8:278-290.

Yildirim I, Freiwald W, Tenenbaum J (2018) Efficient inverse graphics in biological face processing. bioRxiv:282798.

Yin RK (1969) Looking at upside-down faces. Journal of experimental psychology 81:141.

Young AW, Hellawell D, Hay DC (2013) Configurational information in face perception. Perception 42:1166-1178.

Young MP, Yamane S (1992) Sparse population coding of faces in the inferotemporal cortex. Science 256:1327-1331.

Yovel G, Freiwald WA (2013) Face recognition systems in monkey and human: are they the same thing? F1000prime reports 5.

Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? Trends in cognitive sciences 10:301-308.

Zhang P, Jamison K, Engel S, He B, He S (2011) Binocular rivalry requires visual attention. Neuron 71:362-369.

Zhaoping L (2005) Border Ownership from Intracortical Interactions in Visual Area V2. Neuron 47:143-153.

Zhou H, Friedman HS, Von Der Heydt R (2000) Coding of border ownership in monkey visual cortex. The Journal of Neuroscience 20:6594-6611.