2019

From Face Perception to Individual Recognition: The Missing Link

Sofia Mariana Landi

FROM FACE PERCEPTION TO INDIVIDUAL RECOGNITION: THE MISSING LINK

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements

for the degree of Doctor of Philosophy

by

Sofía Mariana Landi

June 2019

FROM FACE PERCEPTION TO INDIVIDUAL RECOGNITION: THE MISSING LINK

Sofía Mariana Landi, Ph.D.
The Rockefeller University 2019

Recognizing other individuals is a key social aspect of our everyday lives. To recognize a familiar individual, we must establish a link between sensory inputs and a representation of that individual held in memory. In primates, faces play a particularly important role on the sensory side of this process, which is reflected in an extensive network of face-selective areas along the inferior temporal lobe. However, where and how memory is re-activated during face perception remains unclear. Using functional magnetic resonance imaging (fMRI), we measured whole brain activity in macaques while they were watching pictures of other monkey faces that were either long-term acquaintances, visually familiar, or totally unfamiliar. In comparison to unfamiliar faces, the entire face-processing network showed increased activity in response to familiar faces of long-time personal acquaintances. In contrast, faces that were only visually familiar elicited less activity than totally unfamiliar faces in most face-selective areas. The face-processing network thus distinguished personally familiar faces from visually familiar faces. Personally familiar faces also prompted the activation of two previously unknown face-selective areas in the temporal lobe. One area was located in the perirhinal cortex (PR), which has been associated with declarative memory, and the other area was embedded in the temporal pole (TP), a region previously associated with social knowledge. These two novel face areas showed a non-linear response as blurred faces became gradually visible, rapidly becoming active when the faces of personal acquaintances became recognizable. Thus, mimicking

the perception of a face approaching us, this paradigm revealed a neural correlate of the 'aha!' recognition moment in face areas TP and PR.

As a first step towards advancing our understanding of the neuronal processing of individual recognition, our fMRI experiments identified two novel face areas specifically involved in recognizing familiar faces. However, the hemodynamic response cannot directly assess neurophysiological properties. Using fMRI-guided electrophysiology, we investigated the responses of neurons within the novel face area TP in awake monkeys, and we provided the first systematic evidence of cells selective for familiar faces. A high fraction of neurons in face area TP were selective for familiar monkey faces, and unfamiliar faces that were physically similar failed to elicit the same neural responses. Importantly, neurons in face area AM, which is thought to compute facial identity at the top of the face perception hierarchy, were not modulated by familiarity. Within TP, neurons also responded to monkey bodies, and to monkey vocalizations. Maximum activity was elicited by the joint observation of faces and bodies, and audiovisual interactions were evident in some TP neurons. Together, these results reveal neuronal processes underlying memory re-activation during face perception and generate hypotheses for testing how individual recognition is achieved through different modalities, thus advancing our understanding into how unique representations of familiar individuals are developed at the neural level.

A mi papá

Emboscado en mi escritura
cantas en mi poema.
Rehén de tu dulce voz
petrificada en mi memoria.
Pájaro asido a su fuga.
Aire tatuado por un ausente.
Reloj que late conmigo
para que nunca despierte.

— *Tu voz,* Alejandra Pizarnik

Acknowledgments

First, I would like to start off by acknowledging that I would NOT be where I am now or who I am without these people in my academic and personal life.

First and foremost, I would like to thank my advisor, Dr. Winrich Freiwald, who has always encouraged me to step outside of my comfort zone. For thinking highly enough of me and pushing me to keep going, to go the extra mile, to not settle.

A big thanks to all the people in Winrich's lab. The ones from the past, the ones from the present. A group of incredible scientists that provided me intellectual and experimental support through these years.

To Alejandra and Lihong, for their endless patience, kindness, and emotional support, and also for their 'invisible' tremendous amount of work that allows research to go on.

To Wilbert, my sensei in and outside of the lab. To him, Julia and Alejandra, for teaching me the ropes in the NHP world. To Ilaria, whose life-wisdom I have been willing to learn since I first met her. To Yuri and Pooja, for being both inspiring and encouraging. To all of them, thanks for being so much more than just lab mates.

To Farid, Geena, and Pooja for helping me edit this document.

A special thanks to Pooja Viswanathan, who contributed to this work and continues with it, and to Belen Zanoni, who also contributed and with whom I learned that mentoring is one of the best parts of doing science.

Outside of the Freiwald lab, many other scientists have lent me vital support along this path. Thanks to my advisory committee – Drs. Vanessa Ruta and Charles Gilbert – for providing me with critiques, advice and support, and for having carefully reviewed each step

of my scientific endeavors at the Rockefeller University. Special thanks to Dr. Nancy Kanwisher, for lending her time and expertise as my external examiner.

To my friends, the ones that are close, the ones that are far, the ones I have known for decades, and the ones I have only got to know deeper in these last years. I am so grateful to the amazing people I met here that made me feel at home. First, to my dance friends, for being my ground wire and with whom I found the best creative outlet: Belu, Mike, KimD, Cristina, Marc, KimS, and Grant. Thanks to the psychopunks, with whom I had so much fun in the last couple of years; specially to Nati and Romain, thanks for accompanying me in most of my thesis writing; and to Alvarito, for helping me put life in perspective. Finally, many thanks to the friends that made me feel part of their family: Tiago, Patricia and Malu; and Christina, Dennis, and Oscar. I am so grateful to the friendship with my ex-roommates: Monica, Pat, Bailey and The Stallion; for making me feel at real home and for the continuous support even after they moved out of the city. To David, for the years of happiness and immense companionship. A big special thanks to my scientist friends: Ceci, Euge, Jushi, Dardo, Gaby, Franco, Mechi, for their incessant and invaluable friendship but also for cultivating my love in science since we first met. To my even older friends: Juli, Ari, Lu y Anita, thanks for all these years of friendship and unconditional support that neither time nor distance can break.

Eternamente agradecida a mis padres y a mi hermana que, a pesar de no tener idea de en qué me estaba metiendo cuando decidí estudiar Biología, me apoyaron incondicionalmente en cada decisión y a cada paso.

These people guided me through it all. Without them, I would not have experienced nor survived some of the most challenging yet empowering moments of my life.

Table of Contents

List of Figures

List of Tables

Chapter I.

Introduction

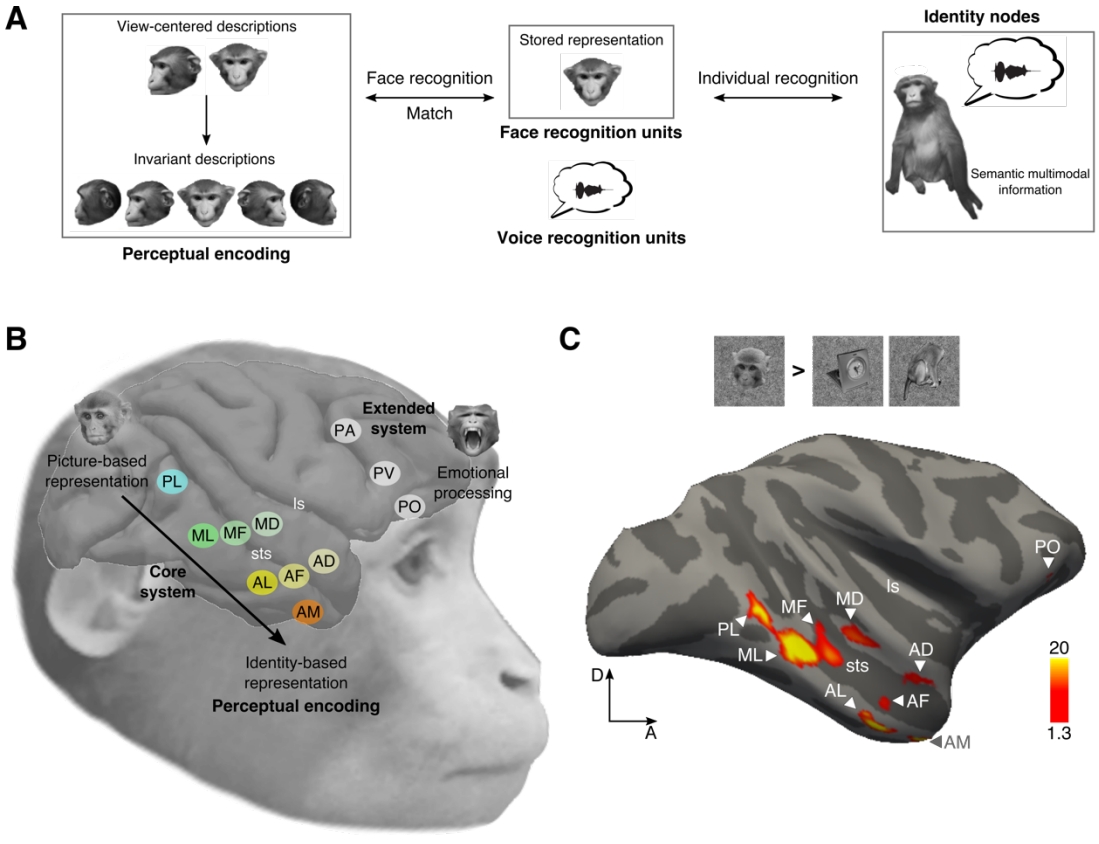
It usually takes less than a blink of an eye to recognize individuals we know, even under strikingly different conditions. For each friend, relative or loyal pet, a unique combination of idiosyncratic traits such as their face, body shape and voice are represented at the cognitive level. Attributes such as a friend's face are represented in regions that analyze visual form and color; their voice, in auditory regions; their manner of dancing, represented in or near the brain regions that respond to the perception of this kind of movement; and so on. Associations between different pairs of attributes (such as faces and voices) could all be processed by a common set of neurons and synapses in a shared and multimodal 'hub' (*1-3*). Each time we recognize someone we know, our brain matches perceptual inputs with this complex and multimodal stored cognitive representation. Faces play a key role in this process. We can easily close our eyes and form a mental image of a familiar face, and even just a glimpse is enough to retrieve certain memories. So how does the brain achieve individual recognition from face perception?

Investigating individual recognition has been heavily influenced by cognitive and neuroscientific models of face and person recognition. Most of these models were built on the highly-influential and well-known model of face recognition proposed by Bruce and Young (1986, Figure 1A). In this model, an initial phase of perceptual encoding produces a set of descriptions of the physical appearance of a face. Recognition of familiar faces involves a match between the products of perceptual encoding and a stored code describing

one of the many faces known to us. Information coming from modality specific "face" recognition units converge onto multimodal hubs or identity nodes (IN), which provide a modality-free gateway to semantic knowledge. This model has been extended to include voice recognition units (*2*, *4-6*) that also converge onto these INs. Through the activation of these nodes, the perception of faces and voices can lead to the spontaneous retrieval of semantic knowledge about familiar individuals, such as their occupation or their nationality (*7*).

Figure 1. Model for individual recognition and organization of the face processing system in the rhesus macaque monkey.

*(**A**) Model of face-identity recognition adapted from Bruce and Young (1986), complemented with a similar path for voice recognition. Processing begins with the generation of a view-centered representation of the face, and structural encoding produces an abstract facial representation that allows for recognition across changes in pose, expression and illumination. Representations of face structure are compared with stored face-recognition units (FRUs), and a match between a given representation and a stored FRU results in the activation of semantic information about the individual in identity nodes (ID). (**B**) Schematic of the macaque face processing system on the pial surface of a right hemisphere (*8*): Core network: areas PL (posterior lateral), ML (middle lateral), MF (middle fundus), MD (middle dorsal, (*9*)), AL (anterior lateral), AF (anterior fundus), AD (anterior dorsal, (*10*)) in the superior temporal sulcus (sts), the anterior medial face area (AM) on the ventral surface of the temporal lobe. Extended prefrontal network (*11*): area PO (prefrontal orbital), in the lateral orbital sulcus, PA (prefrontal arcuate) and PV (prefrontal ventral, corresponding to PL in (*11*)). Prefrontal areas have been found to be modulated by facial expression (*11*). ls: lateral sulcus. (**C**) Definition of face areas: activation maps of the group (fixed-effects) analysis showing regions significantly activated by faces rather than control objects overlaid on the partially inflated right hemisphere of M1's brain. Color-scale indicates negative common logarithm of P value, corrected for multiple comparisons (FDR, $q < 0.05$).*


A
View-centered descriptions
Invariant descriptions
Perceptual encoding
Face recognition
Match
Stored representation
Face recognition units
Voice recognition units
Individual recognition
Identity nodes
Semantic multimodal information
B
Extended system
PA
PV
PO
Emotional processing
Picture-based representation
PL
ls
ML
MF
MD
sts
AD
AL
AF
AM
Core system
Identity-based representation
Perceptual encoding
C
>
PO
ls
MF
MD
PL
ML
sts
AD
AL
AF
AM
D
A
20
1.3

Bruce and Young (1986) did not address the neuroanatomical basis for their cognitive model, probably because lesion studies and electrophysiology were the only sources of information at the time about the neural basis of cognition (*6*). Since then, advancements in the field of face recognition have been remarkable. From the discovery of neurons that responded specifically to images of faces in the 1970s (*12*), to the advent of functional neuroimaging in the 1990s, the field of face processing has found compelling evidence that face perception depends upon specific, anatomically-defined brain regions (*13*) in the inferior-temporal (IT) cortex. More recently, combining functional neuroimaging and electrophysiology has made it possible to target, and compare, the coding of faces in functionally-defined, face selective regions of macaque IT (e.g. (*10*, *14-23*)). This research revealed discrete and interconnected functional regions of the cortex filled with face-selective neurons (*18*, *23*, *24*). While these discoveries have contributed to our understanding of how faces are represented in the brain, comparatively little is understood about how the system represents faces of individuals we know, in comparison to faces of individuals we do not know, and faces of individuals we are getting to know.

In most studies, pictures of unfamiliar faces serve as visual stimuli, perhaps to exploit the greatest experimental advantage that the face-processing system has. Because it is specialized to process only one class of complex forms, and because its computational components are spatially segregated, it offers a remarkable opportunity to dissect the neural mechanisms of face form perception. However, the sole use of unfamiliar face stimuli limits activation in brain regions that process emotional, semantic, and episodic memories about an individual (*25*), leaving us with little knowledge of how seeing the face of a very good friend can lead to a rich emotional and mnemonic experience.

Just the sight of a familiar face is enough to activate memories about a specific individual. Long-term memories involve extensive areas within the medial temporal lobe (MTL, (*26*)). Through learning, nearby neurons in the monkey MTL, specifically in the perirhinal cortex, often respond to physically dissimilar memorized objects (*27*). In the human MTL, a remarkable set of neurons are selectively activated by visually different pictures of the same individuals (person concept cells). In some cases these cells are even activated by letter strings with the individuals' names (*28*), suggesting a sparse code for long-term and more abstract memories. However, the neural pathway that goes from face perception to the computation of such abstract representations of specific individuals is still unknown.

Considering the importance of understanding the interaction between perceptual and memory processes, and the extensive characterization and flexibility in the experimental approaches that the macaque face processing system provides, we decided to explore the impact of familiarity on face processing in this model. This approach would bridge the gap between the neural mechanisms of face perception and long-term memory, thereby bringing us closer to an understanding of how individual recognition can be achieved with just a glimpse of a familiar face.

Face processing in the primate brain

Macaque monkeys live in complex societies with intricate hierarchies. In such a social environment, identifying other individuals reliably is crucial. Studies of monkey temporal lobes have reported scattered single neurons exhibiting strong responses to faces (relative to other categories of objects) (*12*, *29-34*). Following the development of monkey

functional magnetic resonance imaging (fMRI, (*35*, *36*)) several face-selective areas, or "face areas", were identified as regions that respond more to images of faces than objects (*37*). By combining fMRI and single-cell recordings, Tsao et al. (2006) revealed that 97% of the neurons located within an fMRI-defined face area in the macaque brain are selective to faces. Similar proportions have been found in other face areas in the macaque brain (*18*), suggesting that macaque face areas constitute a specific level of functional organization: discrete areas dedicated to face processing.

A fixed number of face-selective areas are located in the temporal lobe. These 'core' areas are highly reproducible across studies and stereotyped across individuals (*8*) (Figure 1B-C). Three of these areas sit along the ventrolateral lip of the superior temporal sulcus (STS), running from its middle to its anterior end: the posterior lateral face patch (PL), the middle lateral face patch (ML), and the anterior lateral face patch (AL). Two additional areas lie in the fundus and upper bank of the STS: the middle fundus face patch (MF) and the anterior fundus face patch (AF). The most anterior face area (AM, anterior medial) is found on the ventral surface of the anterior temporal lobe near the anterior middle temporal sulcus (AMTS). Both electrical micro-stimulation during fMRI (*24*) and anatomical tracer studies (*38*) revealed that face areas are selectively interconnected with each other. After these initial discoveries, two new face areas that were also highly reproducible across individuals were discovered in the dorsal bank of the STS: middle dorsal (MD, (*9*)) and anterior dorsal (AD, (*10*)). While their functional and anatomical connectivity to other face areas remains undetermined, we will assume that given their reproducibility they form part of the core network of face areas. This network offers an

extraordinary and unique preparation to dissect the neural mechanisms underlying perception of a circumscribed set of complex forms.

At the single neuron level, the spatial segregation of face-selective areas makes it an ideal system to study how visual responses to face forms evolves across different components of the network. Electrophysiological experiments show not only that all IT face areas contain high concentrations of face cells (*14*, *16*, *18*, *23*), but also that the different areas are functionally distinct. The most posterior area PL has been shown to be selective to the eye region of the face (*14*). Neurons in the middle areas MF and ML respond to specific views and are broadly tuned to identity. Cells in AL tend to respond to mirror-symmetric views and are more narrowly tuned to identity than MF and ML. Some neurons in AM are even more narrowly tuned to identity and exhibit nearly complete viewpoint invariance. This functional organization makes it possible to explore the impact that familiarity might have on face representation along the hierarchy, from early picture-based in the posterior face areas into an identity-based representation (*33*, *39*) in the anteromedial face area AM (*18*) (Figure 1B-C).

Beyond the temporal lobe, regions of face-selectivity have also been found in the prefrontal cortex (*11*, *40-42*): area PO (prefrontal orbital), in the lateral orbital sulcus, PA (prefrontal arcuate), and PV (prefrontal ventral, corresponding to PL in (*11*)). Though distant from the IT and early visual cortex, prefrontal areas have been found to be modulated by facial expression (*11*) and other social categories (*43*), making them also sites of interest to investigate familiarity effects.

Several properties of the macaque face-processing system make it an ideal model to study how face perception leads to the recognition of familiar individuals. First, it allows

us to build on the detailed knowledge of the macaque core face-processing system and its physiology, not available in humans or any other species. Second, the uniquely high consistency of brain morphology and face network topology in the rhesus monkey provides the necessary foundation for looking into the neural architecture supporting the recognition of familiar faces. And finally, we hoped that the experimental accessibility to perform both fMRI and electrophysiological recordings from the same subjects would allow us to determine the detailed neural code of familiar face processing in comparison with neural codes in other face areas. Beyond these general benefits, the known differences between familiar and unfamiliar face recognition make faces the ideal object to explore how individual recognition is achieved from perceptual inputs.

The special status of familiar faces

Most of what we know about how primates recognize familiar individuals comes from human experiments. Humans recognize personally familiar faces exceptionally well. This is true despite large variations in lighting, viewpoint, facial expressions, or disguises, such as hats or glasses. In fact, human recognition of unfamiliar faces is remarkably bad (*44-46*), with sometimes consequential effects, as e.g. for eyewitness identification. Familiar face recognition is also not just a matter of visual appearance: retrieval of person knowledge and emotional responses to personally familiar faces add to their unusual significance (*25*, *47*). While human face processing abilities are most remarkable for familiar faces, the majority of empirical evidence regarding its neural basis stems from experiments involving unfamiliar face stimuli. Thus, we know little about the neural basis of how face perception leads to the recognition of familiar individuals.

The role of familiarity in face processing is still unclear. Both neuroimaging and patient studies have pointed to differential recruitment of common neural machinery like the fusiform face area (FFA, (*13*)) and to the involvement of a number of various additional brain areas in the frontal and anterior temporal lobes (*48-50*). Neuroimaging studies show that there is a progressive disentangling of the representations of face identity from face view towards both frontal lobe regions (*51*, *52*) and the anterior temporal lobe regions (*53-55*). Several patient studies report lesions in the anterior temporal lobe associated with face-identity recognition deficits, and also multimodal impairments, for example, impaired recognition of both faces and voices (for a thorough review see (*50*)). Yet because of the variability of findings across studies (*49*, *50*, *54*), unspecific task activations (*56*), and technical difficulties imaging parts of the architectonically complex temporal lobes (*57*), among other factors, it has remained difficult to draw firm conclusions about the location of familiar face recognition processes. Furthermore, because different functions were assigned to similar regions, and both localized and distributed face representations have been found (*53*), the functional specificity and functional organization of the regions processing familiar faces has remained unclear as well.

Human studies have provided only provisional indication of regions where familiar face recognition could take place. The macaque face-processing system provides an appealing experimental alternative to deepen our understanding of the neural mechanisms at work: a range of experimental techniques are feasible, allowing for the temporal and spatial resolution necessary to probe single-cell and small network activity. However, in macaques, even less is known about the neural systems supporting familiar face recognition.

Rhesus monkeys can recognize familiar individuals from static pictures and voices (*58*). Like humans, monkeys' performance in unfamiliar face recognition is affected by changes in viewpoint: when there's a change from frontal to other head views, performance drops from 90% to 20% for unfamiliar faces, while for personally familiar faces performance remains intact (*59*). Only a few electrophysiology studies targeting the anterior and middle temporal lobe have used personally familiar faces as stimuli in macaques (*59-64*). Young and Yamane (1992) and Eifuku et al. (2011) reported neural population codes reflecting the level of familiarity with human faces (*60*, *61*) in the upper bank of the most anterior portion of the STS and in the anterior IT respectively. Sliwa et al. (2014) and Munuera et al. (2018) recorded from regions in the MTL. Sliwa et al. (2014) found that neurons in the hippocampus respond differentially to social stimuli, but are not affected by familiarity. Munuera et al. (2018) showed that the amygdala encodes the social hierarchical rank of familiar individuals. Finally, Nakamura et al. (1994) recorded from a large region that included the temporal pole, which is the most anterior part of the temporal lobe. The neurons recorded in the temporal pole were not face-selective, however, some responses to specific familiar faces and objects were reported. Perhaps because none of these studies tested face-selective cells specifically, and although some of them found familiarity effects at the population level, none of them showed a systematic difference in the responses of individual neurons to familiar and unfamiliar faces.

Although one of the goals of face processing in social species is the recognition of familiar individuals, little is known about the neural systems supporting familiar and unfamiliar face recognition. Several questions remain unanswered: (1) Do familiar and unfamiliar face recognition use the same neural machinery but with different efficiency?

The more we encounter an initially unfamiliar face, the more familiar it becomes, and the more existing circuitry might become tuned to it. Differential use of the same circuitry would explain this quantitative difference between familiar and unfamiliar face recognition. If this is the case, (2) how are face-selective neurons in functionally defined face areas modulated by familiarity? And (3) is this modulation the same across areas, or, for example, are the more anterior face-selective areas more sensitive to familiarity? Given the mentioned qualitative differences between familiar and unfamiliar face recognition, (4) do familiar and unfamiliar faces use different neural systems, each implementing a different computational strategy to process them? Finally, going beyond face perception, (5) do other identity features such as body shape, or vocalizations modulate the activity of these systems?

Experimental Strategy

To address the questions raised above and identify brain regions supporting the representation of familiar faces in the macaque, we ran a set of fMRI experiments first. We imaged the whole brain of awake macaques while showing them still pictures of personally familiar, visually familiar and unfamiliar faces and non-face objects. This allowed us to carefully compare general face selectivity, responses to personal and visual familiarity in the whole brain, and make inter-areal comparisons in an unbiased manner. In order to examine regions where face familiarity might be critical for neural activity, we located an extended face processing system by comparing responses to all faces with responses to all objects. Afterwards, we considered the activity of each face-selective area when a monkey viewed pictures of familiar and unfamiliar faces coming into focus in order to determine

which areas – if any – demonstrated an instantaneous nonlinear effect of individual recognition. Through this experiment, we found two previously undiscovered face-selective areas with a functional specialization for familiar faces.

Our fMRI experiment posited two brain regions where the interaction between face processing and individual recognition might take place at the single-cell level. Using fMRI guided electrophysiology, we studied the activity of single cells in one of these areas and compared it to the activity of cells in a face-selective area in the core face-processing system.

These two studies, by identifying areas that process familiar faces, and by characterizing one of these areas, have deepened our understanding of the brain's face-processing machinery and have set the stage for future exploration of individual recognition neural foundations.

Chapter II.

The layout of neural circuits for familiar face recognition

To explore at a systems level the neural basis of familiar face recognition, we first localized the face-processing system in four rhesus monkeys with whole brain fMRI (*Face Localizer Experiment,* see *Experimental procedures*) by computing the contrast between neural activations to unfamiliar monkey faces versus objects. This contrast revealed the known stereotypic face-processing areas in all subjects (see Figure 1B, controlling the false discovery rate -FDR- at $q<0.05$ (*65*)). We took advantage of this organizational specificity to determine the functional organization of systems for familiar and unfamiliar face processing. We asked the following questions (Figure 2A): Do familiar and unfamiliar faces recruit the same face processing networks? If so, does familiar face processing engage one or several nodes differently? Do familiar faces engage the entire system more efficiently? And do effects of familiarity grow stronger as representations become more identity selective? Because anterior inferotemporal cortex has been suggested as a location for familiar face processing in humans (*66*), the most anterior face areas in the macaque face processing system might be particularly selective for familiar faces (i.e. face area AM). Alternatively, does familiar face processing rely on additional machinery outside these systems, and if so, where is it located?

Past neuroimaging studies in macaque monkeys have used pictures of faces unfamiliar to the subjects. However, unfamiliar face pictures become visually familiarized

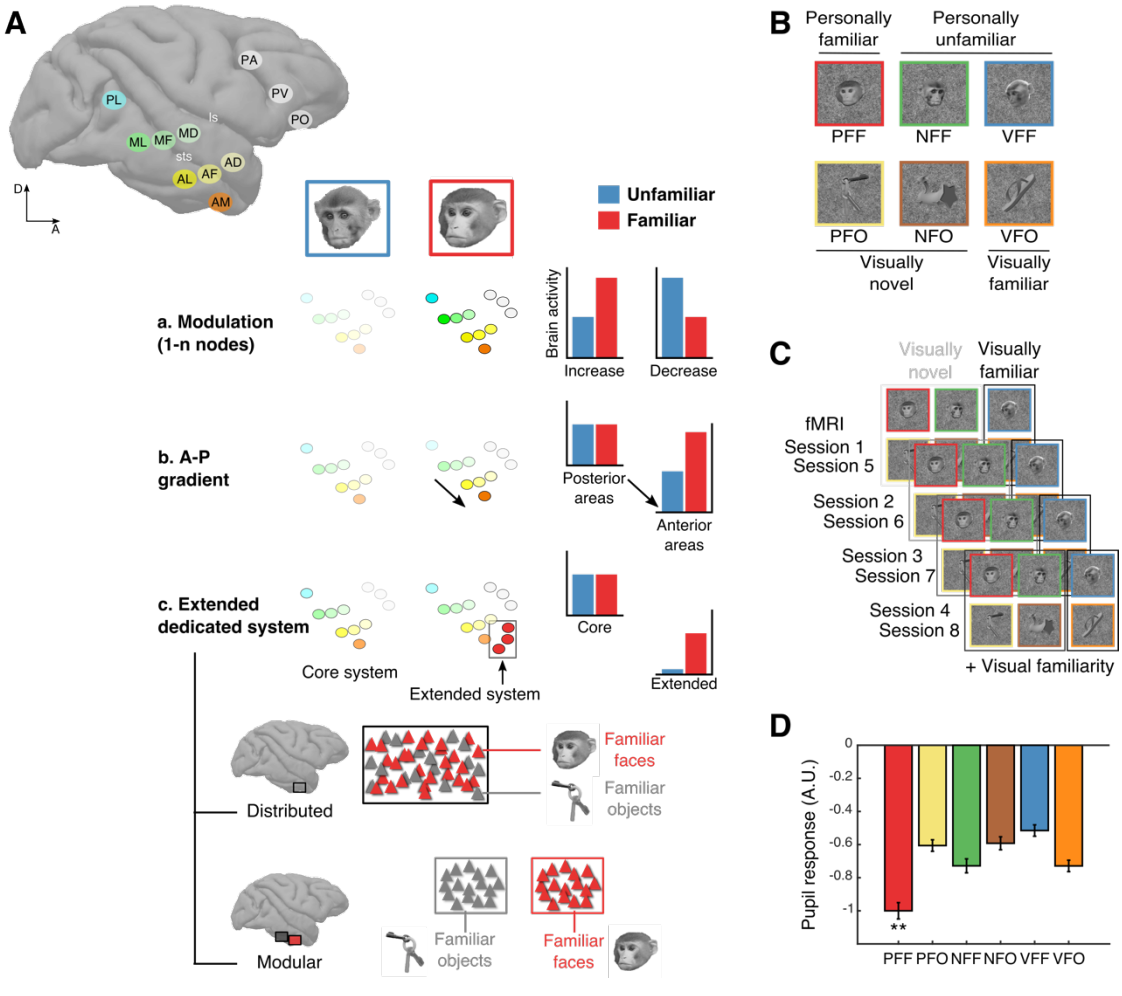
during the frequent and repeated exposures typical for monkey imaging studies. The nature of familiarity effects can thus stem from real life (i.e. personal familiarity) or repeated exposures to stimuli in experimental settings (i.e. visual familiarity). Visual and personal familiarities differ by many psychologically relevant factors, such as real-world character, variety and amount of exposure, social knowledge, and emotional relevance. To ensure recognition, pictures of personally familiar faces (PFF) were taken of four subjects which had been housed together for over two years before the beginning of the experiments. Pictures of personally familiar objects (PFO) were taken of toys with which the subjects interact daily (see Figure 2B and *Experimental procedures*). This minimized two typical problems of stimulus design in the study of familiarity: 1) inter-subject variability in the degree of familiarity, and 2) the picture-specific nature of visually familiarized stimuli (*67*, *68*). To assess the nature of familiarity effects, we also used pictures of personally unfamiliar rhesus monkeys, and made half of them visually highly familiar (visually familiar faces, VFF, which were seen hundreds of times by the subjects before the experiments, see *Experimental procedures* for details), while the others were completely unfamiliar (non-familiar faces, NFF). We generated analogous stimulus categories for objects: visually familiar objects (VFO, seen as much as VFF) and non-familiar objects (NFO), and defined an object-selective area to explore familiarity effects for these categories (see *Experimental procedures* and Figure 2B). All these stimuli became visually familiar over the course of four fMRI sessions. On session five we changed stimuli such that PFF, PFO, NFF and NFO were visually novel again. While the identity of faces and objects remained the same, their pictures were taken from different viewpoints (e.g. a face that was shown in frontal view during the first four sessions was shown in profile view in

the last four sessions). VFF and VFO remained unaltered throughout all sessions (see *Experimental procedures* and Figure 2C).

Given a rich past literature that shows that rhesus macaques can identify faces from static pictures and personally familiar individuals from static pictures of their faces (*58*, *69-73*), we did not conduct a behavioral experiment to test whether the PFF were indeed recognized by the monkeys as belonging to a personally familiar individual. However, we did confirm that all four individuals showed a greater pupillary constriction during perception of pictures of personally familiar faces compared to the other conditions (one-way ANOVA with stimulus category as factor, $F_{(5,2448)} = 9.8$, $p<10^{-9}$, post hoc tests using Tukey's honest significant difference test $p< 0.01$). This result is in line with studies that show that in humans, pupil constriction is greater in response to upright conspecific faces compared to other primates' (less familiar) faces (*74*), and with studies in macaques that show a greater pupil constriction is observed in response to social images compared to non-social images (*75*).

Figure 2. Familiar face processing in the non-human primate: hypotheses, experimental design, and behavior.

*(**A**) Left: schematic of the macaque face processing system on the pial surface of a right hemisphere, for details see Figure 1A. Below, hypothetical scenarios for familiar face processing. Differential responses are depicted in a darker tint. (a) There is a modulation by familiarity in the classical face processing system. It may either increase or decrease activity. (b) There is an anterior-posterior gradient of modulation, with identity-selective representations particularly selective for familiar faces. Activity increases with familiarity in the more anterior face areas, as other qualities do (*18, 76*). (c) Familiar face processing relies on additional brain areas outside the core face processing system. While the core system does not differentiate between familiar and unfamiliar faces, an extended face system exists that is highly selective for familiar faces. The extended system could code for face familiarity in (a) a distributed or (b) a modular manner. (**B**) Example stimuli: pictures of personally familiar faces and objects (PFF and PFO) and pictures of non-familiar faces and objects (NFF and NFO) were visually novel to the subjects, while faces and objects that were seen hundreds of times before the experiment were considered visually familiar (VFF and VFO, see Experimental procedures). (**C**) Experimental design. PFF, NFF, PFO, and NFO were initially visually novel and became visually familiar over the course of first four days of scanning. On day five these stimuli were changed and were thus visually novel again. VFF and VFO were not changed over the course of the experiments. (**D**) Average normalized pupil responses of four subjects (left to right) across six different stimulus conditions of the Face and Object Familiarity Experiment: PFF (red), PFO (mid-level gray), NFF (green), NFO (dark gray), VFF (blue) and VFO (light gray). **p<0.01, corrected using Tukey's honest significant difference test. Error bars represent standard error of the mean (SEM).*


A
PA
PV
PL
PO
ls
ML
MF
MD
sts
AD
AL
AF
AM
D
A
Unfamiliar
Familiar
a. Modulation
(1-n nodes)
Brain activity
Increase
Decrease
b. A-P
gradient
Posterior
areas
Anterior
areas
c. Extended
dedicated system
Core
Core system
Extended system
Extended
Familiar
faces
Familiar
objects
Distributed
Modular
Familiar
objects
Familiar
faces
B
Personally
familiar
Personally
unfamiliar
PFF
NFF
VFF
PFO
NFO
VFO
Visually
novel
Visually
familiar
C
Visually
novel
Visually
familiar
fMRI
Session 1
Session 5
Session 2
Session 6
Session 3
Session 7
Session 4
Session 8
+ Visual familiarity
D
Pupil response (A.U.)
0
-0.2
-0.4
-0.6
-0.8
-1
**
PFF PFO NFF NFO VFF VFO

Whole-brain fMRI: personally familiar faces reveal two novel face-selective areas in the temporal lobe

All temporal and prefrontal face areas (mapped with the independent Face Localizer Experiment, see *Experimental procedures*) were activated more by PFF than by PFO (Figure 3A top panel, FDR corrected at $q<0.05$, see *Experimental procedures* and Figure 4). In addition, PFF recruited two discrete, previously unrecognized face areas in the anterior temporal lobe: one in perirhinal cortex (PR, (*77*)) and one in the temporal pole (TP, (*78*)) (Figure 3A lower panel, contrast PFF vs. PFO, FDR corrected at $q<0.05$).

The two novel areas were present in both hemispheres of all four subjects (Figure 3B). They were located anterio-medially to face area AM, so far considered the top of the face-processing hierarchy (*18*). TP was located at the anterior end of the temporal pole in area TGsts (*78*) (Figure 3A, B and D) while PR was located in the rostrolateral subdivision of perirhinal cortex, area 36rl (Figure 3A & D) (*77*). TP, PR, and AM are located in regions TGsts, 36rl, and TEav, which are all reciprocally interconnected (*78*, *79*). Locations and sizes of TP and PR were highly conserved across subjects and hemispheres (Figure 3B, Tables 1 and 2). Reliable occurrence and consistent topography suggest areas TP and PR constitute two new face areas extending the previously known 'core' system, probably through connections with AM, deeper into the temporal lobe.

Figure 3. Personally familiar faces recruit the core face processing system as well as two areas in perirhinal cortex and the temporal pole.

*(**A**) Top panel: inflated hemispheres of M1 showing the regions that responded significantly more to PFF than to PFO in the group (fixed-effects) analysis (FDR corrected at q = 0.05). Color-scale indicates negative common logarithm of p-value. Lower panel: results are overlaid on coronal slices of the average template brain (see Experimental procedures). Relative slice position is shown in the top panel (dotted lines), and the anterior/posterior position is indicated at each slice's the top right corner (mm relative to the interaural canal). ls: lateral sulcus, sts: superior temporal sulcus, amts: anterior middle temporal sulcus, rs: rhinal sulcus. (**B**) Coronal slices for the four subjects (M1-4) showing positions of TP, PR, and AM for the contrast PFF, VFF, NFF > PFO, VFO, NFO. The anterior/posterior position of each slice is indicated in the top left corner (mm relative to the interaural canal). (**C**) Object selective area, approximately 4 mm posterior to the face area AL (*80*). Inflated hemispheres of M1 showing the regions that responded significantly more to PFO than to PFF in the group (fixed-effects) analysis (FDR corrected at q = 0.05). (**D**) Pial surface of monkey M2 showing the position of TP and PR, as well as other face areas for comparison. (**E**) Contrast effect sizes to VFF vs. NFF (red) and VFO vs. NFO (gray) relative to the face selectivity of each ROI. *p<0.05, **p<0.01, ***p<0.001, corrected using Holm-Bonferroni Experimental procedures. Error bars represent standard deviation. (**F**) Same as D but for PFF vs. NFF and PFO vs. NFO. (**G**) Contrast effect sizes (PFF vs. NFF, VFF vs. NFF, PFO vs. NFO, VFO vs. NFO) for the four grouped ROIs were analyzed with a 2 (stimulus type: face/object) x 2 (familiarity type: personal/visual) ANOVA. The interaction of stimulus and familiarity type influenced contrast effects in the four grouped ROIs (p < 0.05, corrected for four multiple comparisons using FDR; for temporal core face areas F(1,124) = 4.75, for the new temporal lobe face areas F(1,28) = 4.87, for the prefrontal face areas F(1,44) = 7.18, and for the object-selective area F(1,12) = 10.33). Post hoc tests are shown with asterisks: *p<0.05, **p<0. 01, ***p<0.001, corrected using Tukey's honest significance difference. Comparisons marked 'ns' did not reach significance.*

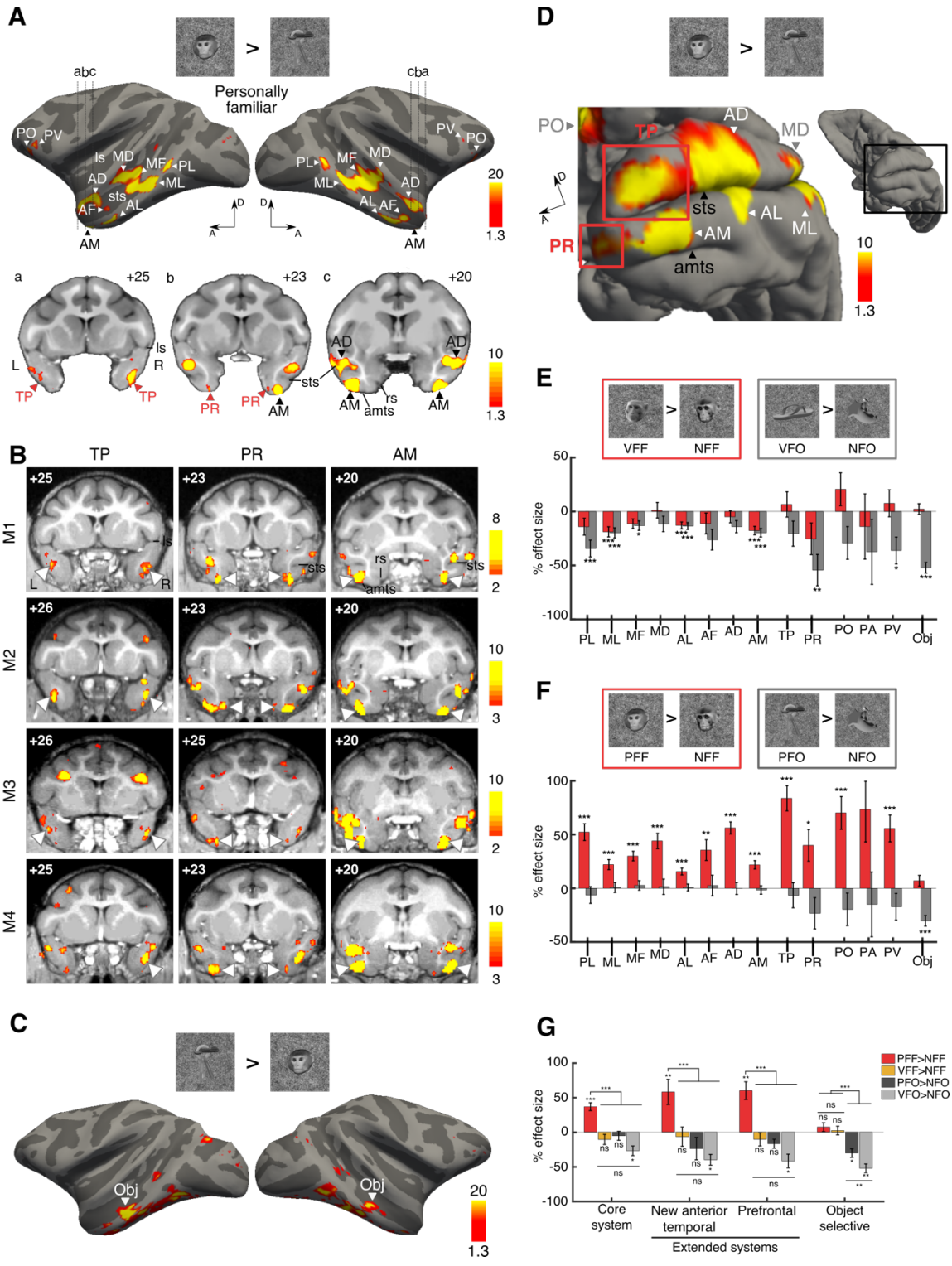

A
abc
cba
Personally familiar
PO PV
ls MD MF PL
AD ML
sts
AF AL
AM
PV PO
PL MF MD
ML AD
AL AF
AM
20
1.3
a +25
b +23
c +20
L
R
ls
TP
PR
sts
AM
amts
rs
AD
10
1.3
B
TP
PR
AM
M1
M2
M3
M4
+25
+23
+20
+26
8
2
10
3
C
Obj
20
1.3
D
PO
TP
AD
MD
sts
AL
ML
AM
PR
amts
10
1.3
E
VFF
NFF
VFO
NFO
% effect size
50
0
-50
-100
PL ML MF MD AL AF AD AM TP PR PO PA PV Obj
F
PFF
NFF
PFO
NFO
100
50
0
-50
G
PFF>NFF
VFF>NFF
PFO>NFO
VFO>NFO
% effect size
Core system
New anterior temporal
Prefrontal
Object selective
Extended systems

Table 1. Size of TP, PR, and AM.

	TP		PR		AM	
	LH	**RH**	**LH**	**RH**	**LH**	**RH**
M1	22.0	24.0	19.9	6.1	93.6	37.3
M2	45.6	47.8	19.3	13.0	57.5	17.3
M3	10.3	26.5	7.6	13.5	59.4	91.6
M4	7.0	56.6	23.6	26.4	97.0	55.3
Mean	**21.2**	**38.7**	**17.6**	**14.8**	**76.9**	**50.3**
Std	**17.5**	**16.0**	**6.9**	**8.4**	**21.3**	**31.6**

The size (in mm^3) of TP, PR and AM in the left and right hemispheres of the four macaques tested, at FDR corrected q=0.01.

Table 2. Position of TP, PR, and AM.

	TP						PR						AM					
	LH			RH			LH			RH			LH			RH		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
M1	-19	26	4	18	28	3	-14	25	-2	13	27	-1	-19	22	1	16	23	0
M2	-21	26	1	18	25	2	-13	23	-3	10	23	-2	-20	21	-4	15	21	-4
M3	-21	26	5	19	25	3	-15	24	-1	13	22	-2	-17	22	-2	16	19	0
M4	-17	28	1	20	25	2	-14	25	-3	15	23	-3	-18	21	-2	21	21	-1

*Stereotactic coordinates following conventions in (*81*) of the peak voxel of PR, TP, and AM (left to right) in the left and right hemispheres of the four macaques (top to bottom) tested. X: mm from midline to the right (positive) or left (negative), Y: mm rostral (positive) or caudal (negative) to interaural line, Z: mm dorsal (positive) or ventral (negative) from interaural line.*

Nature of familiarity effects in face areas

We first analyzed the overall effect of visual and personal familiarity as the modulation in activity relative to unfamiliar stimuli (contrasts VFF/O > NFF/O and PFF/O > NFF/O, normalized to face/object selectivity) with a region of interest (ROI) analysis (see *Experimental procedures*) across all sessions. Visual and personal familiarity systematically modulated the activity of all face and object areas (Figure 3C, 3E & F, Figure 4). Visual familiarity with faces and objects led to significantly reduced activity in many, but not all, face and object areas (Figure 3E & F). Personal familiarity with faces, in contrast, enhanced activity in all face areas and groups, while personal familiarity with objects reduced activity in the object area (Figure 3C, 3E-F). To run statistical comparisons, we sorted ROIs in four groups of areas: the temporal core face areas, the new temporal lobe face areas, prefrontal face areas and object selective temporal cortex (Figure 3G). The main effects of familiarity were enhanced activity in response to personal familiarity in face areas (Figure 3G, $p < 0.01$, without significant differences in modulation between the three groups of face areas), and general activity reduction in responses to familiar objects (Figure 3G and Figure 4). In summary, familiarity effects were widespread in face and object selective areas. These effects were strong, enhancing or suppressing, and highly specific depending on local specialization, stimulus category, and the nature of familiarity (Figure 3G, 3-way ANOVA interaction effect, $F(3,48) = 2.79$, $p = 0.05$).

To understand the nature of the familiarity effects, we analyzed how activity evolved over several days of exposure to visually novel faces (PFF and NFF, Figure 3C). During the first fMRI session, visually novel faces elicited stronger activity than visually familiar faces for all face areas ("novelty effect"), with PFF exhibiting the strongest

activation (Figure 4). With repeated presentations of the stimuli, response in all face areas decreased for both PFF and NFF in most face areas (Figure 4). The novelty effect for personally familiar faces decreased over the course of the first four fMRI sessions, and it was already significantly different between the first and second sessions for most face areas (Figure 5A). The novelty effect for NFF was smaller and the difference between sessions was not as strong or significant as with PFF (Figure 5B). Introducing a visually novel set of stimuli in session five restored the novelty effect in all face areas, but only for PFFs (compare sessions 4, 5 and 6 in Figures 4 and 5A). The novelty effect for NFF was observed only in areas ML, MF, AL, AF, and AM (Figure 4 & 5B). Visual familiarity with images of objects had a similar effect in the object-selective area: visually novel objects elicited initially stronger activations than visually familiar objects, however, the effect was stronger for NFO than for PFO (figures 4 and 5C, 5D). In the object-selective area, and in some face areas, repeated presentation decreased the novelty effect for NFO and PFO. Thus, simple visual familiarity modulates activity over face and object selective cortex, weakening the response with repeated exposure.

Figure 4. Time course of familiarity in face areas.

Percent fMRI signal change compared to baseline for VFF (blue circles) and VFO (orange circles), PFF (red squares), PFO (yellow squares), NFF (green diamonds), NFO (brown diamonds), in the ROIs defined by the contrast PFF > PFO of the Face and Object Familiarity Experiment. Data averaged across four monkeys. Error bars indicate 95% confidence intervals; non-overlapping bars indicate statistical significance without correcting for multiple comparisons. See figure 5 for the comparison of effects across fMRI sessions.

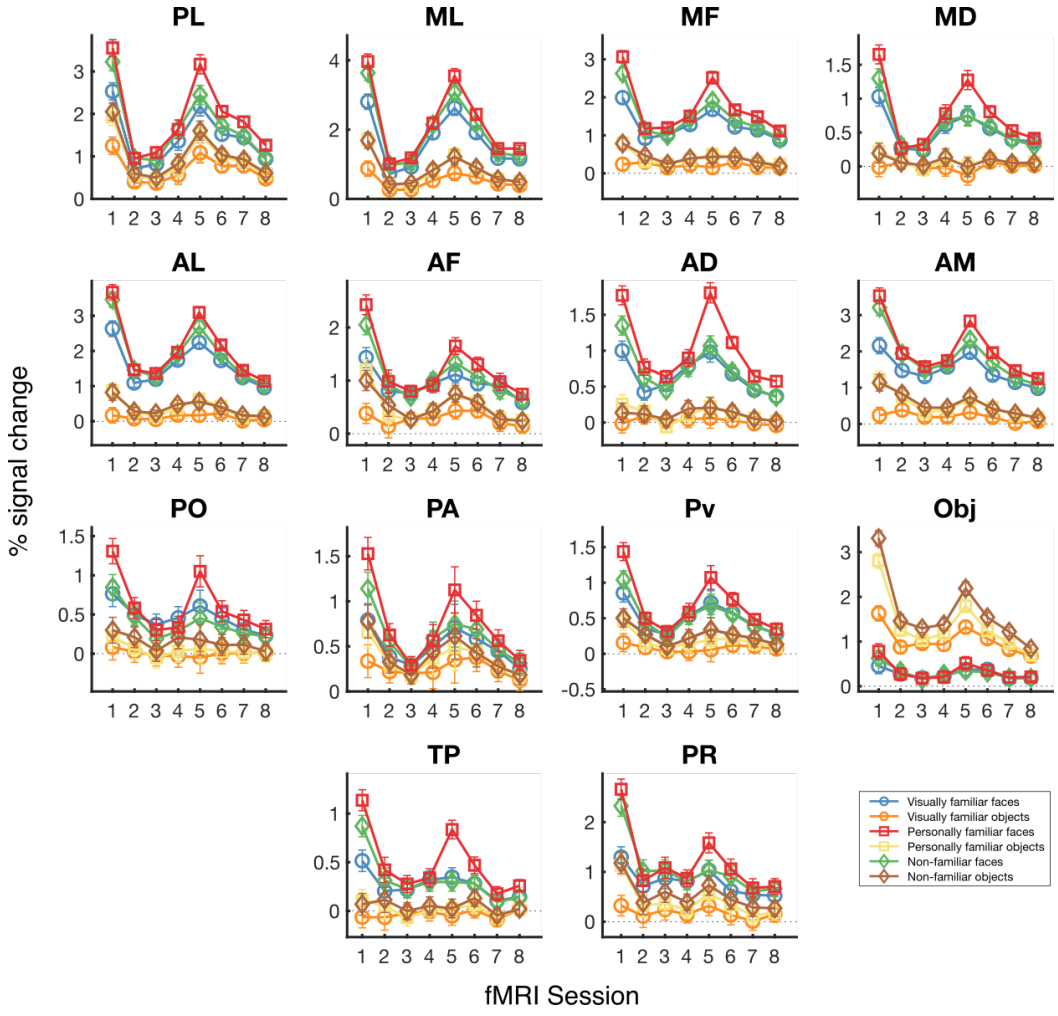
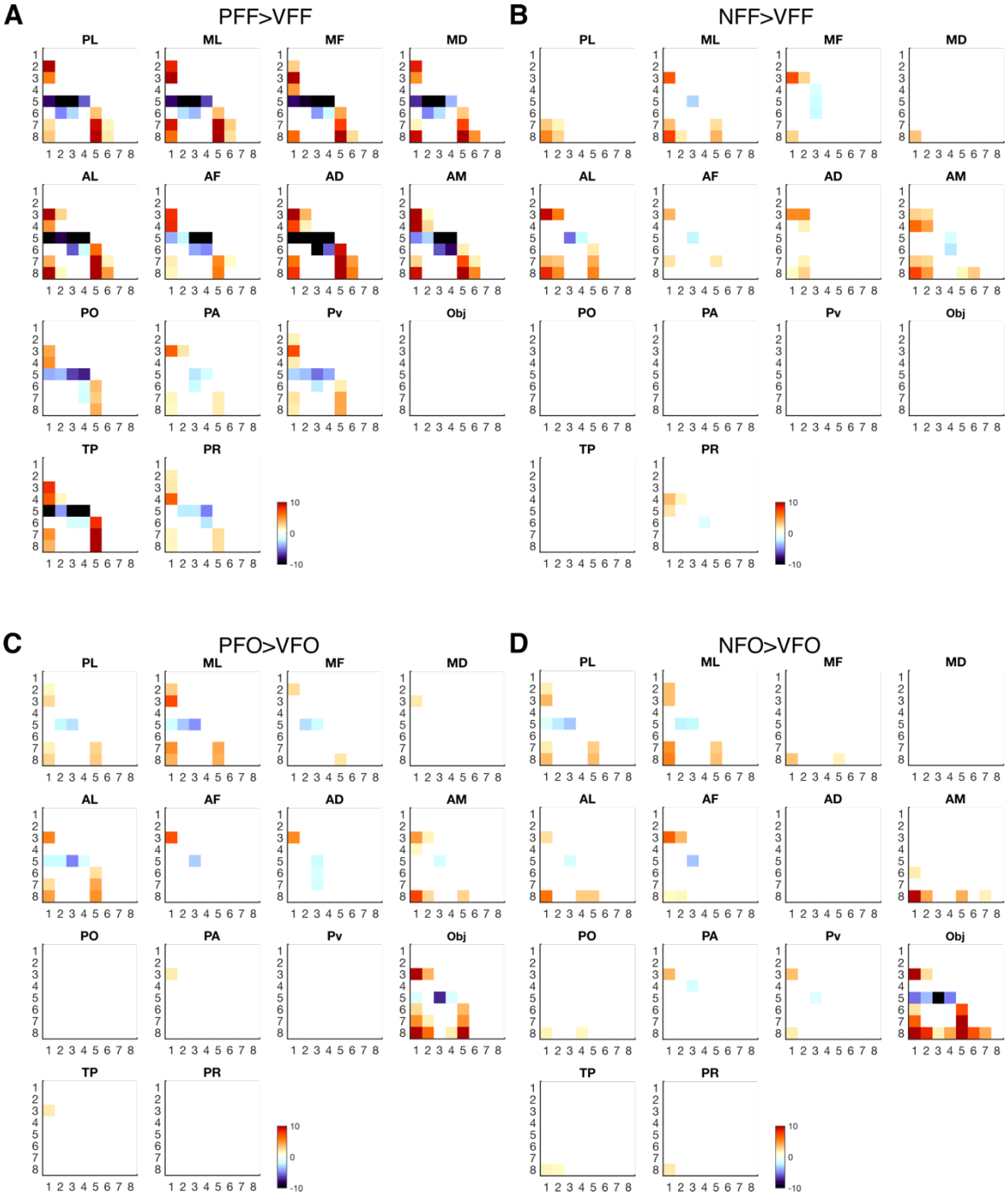

PL
ML
MF
MD
AL
AF
AD
AM
PO
PA
Pv
Obj
TP
PR
% signal change
fMRI Session
Visually familiar faces
Visually familiar objects
Personally familiar faces
Personally familiar objects
Non-familiar faces
Non-familiar objects

Figure 5. Visual familiarity effects over eight fMRI sessions.

*Statistical significance of the difference between fMRI sessions for the contrasts PFF>VFF (**A**), NFF>VFF (**B**), PFO>VFO (**C**), NFO>VFO (**D**). Color scale indicates common logarithm of p-value. Positive (negative) values indicate that a contrast's effect size was higher for the earlier (later) fMRI session in the comparison, corrected using Holm-Bonferroni methods.*


A
PFF>VFF
B
NFF>VFF
C
PFO>VFO
D
NFO>VFO
PL
ML
MF
MD
AL
AF
AD
AM
PO
PA
Pv
Obj
TP
PR
1 2 3 4 5 6 7 8
10
0
-10

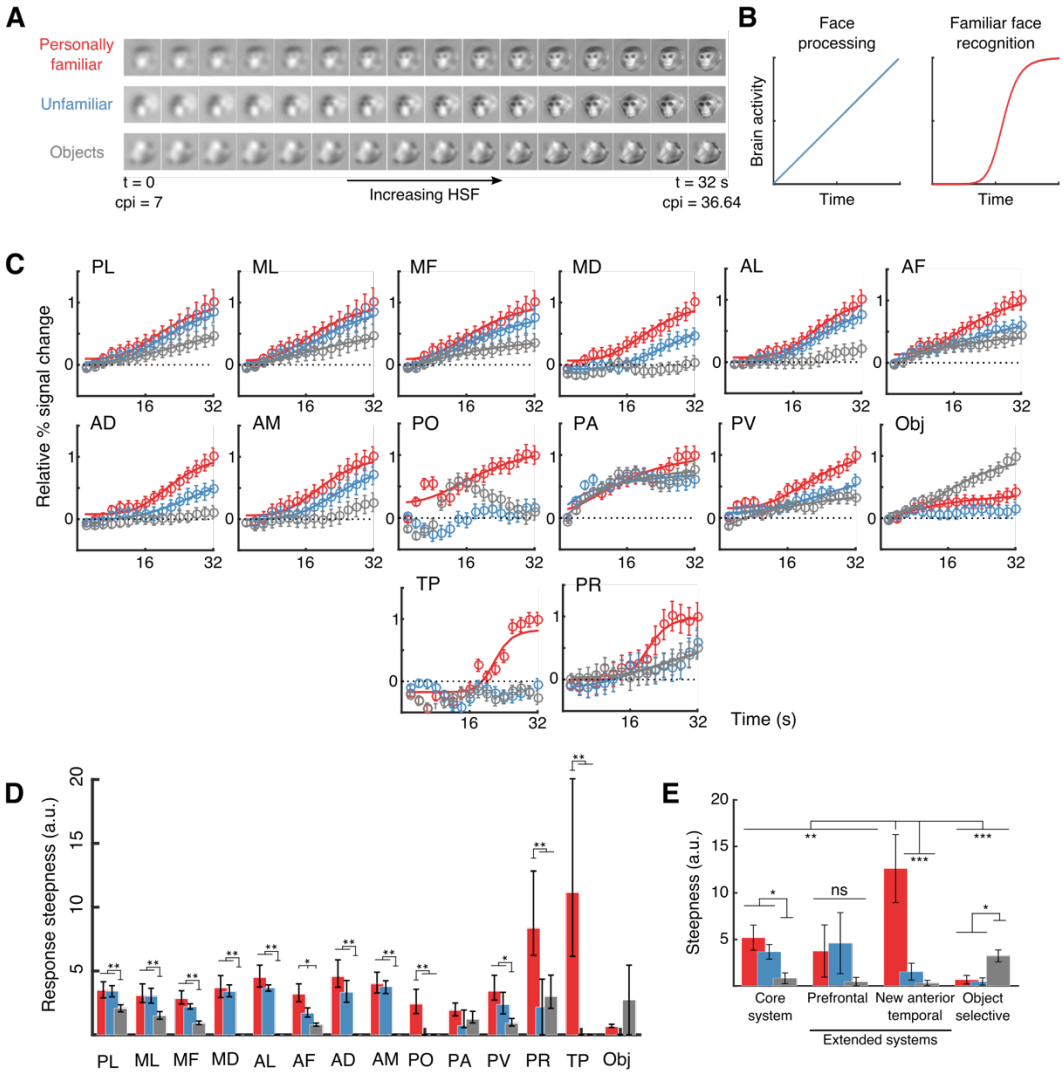
Specific familiarity signature in the temporal pole and the perirhinal cortex

A hallmark of familiar face processing is efficient recognition even during partial occlusion or severe blurring (*82*, *83*). Under these conditions, face information might be processed for a long time, before a sudden transition to recognition. A paradigm sensitive to this signature has recently been introduced (*84*). Here, initially highly blurred and unrecognizable stimuli slowly incorporate, over the course of seconds, increasing amounts of high-spatial frequency (HSF) information (Figure 6A). With this type of stimulation, activity in generic face and object processing systems is expected to increase linearly, in parallel to information accumulation. Instead, familiar face recognition systems are expected to (additionally) nonlinearly accelerate activation upon recognition (*84*) (Figure 6B). We used three sets of stimuli- objects, unfamiliar, and personally familiar faces, which were unblurred over the course of 32 seconds (see *Experimental procedures*). Activity in the core face areas and in prefrontal face area PV ramped up concomitantly, exhibiting a preference for faces over non-face objects and of personally familiar over unfamiliar faces throughout the stimulation period (Figure 6C). Response time courses in prefrontal face area PO differed markedly, exhibiting a face familiarity preference early on which was maintained throughout stimulation. This pattern of results is compatible with the hypothesis that PO utilizes low-spatial frequency information to form a 'first guess' of stimulus identity (*85*). Face areas TP and PR, however, exhibited a highly non-linear response increase, and this accelerated response increment occurred for PFFs only. TP was not even activated by any of the other stimuli (permutation tests $p > 0.1$). To quantify the response non-linearity, we fit response trajectories with a sigmoidal function (Naka-

Rushton function (*86*), Figure 6C, see *Experimental procedures*), whose exponent quantifies response steepness. All core and prefrontal face areas exhibited significant differences in response steepness for faces versus objects (Figure 6D & E), but only anterior temporal face areas TP and PR showed high response steepness and highly significant differences between familiar and unfamiliar faces (Figure 6D & E). Response steepness was significantly higher in the anterior temporal face areas than in the other three functional groups was highly significant (Figure 6E, $p < 0.001$).

Figure 6. TP and PR possess a unique signature for face familiarity.

*(**A**) Stimuli and experimental design: each trial started with a familiar, unfamiliar face or object containing the same spatial frequency content (7 cycles per image, c.p.i., details in Experimental procedures). HSF content then increased progressively over 16 TRs (32 seconds). (**B**) Schematic summarizing the logic of the experimental design (84): activation in areas involved in generic shape processing will increase, approximately linearly, with the revelation of new visual information. In contrast, activation in an area supporting familiar face recognition will remain inactive until the (sudden) occurrence of recognition. **(C)** Stimulus-aligned time courses within face-preferential ROIs and an object-preferring ROI, during the presentation of familiar faces (red), non-familiar faces (blue), and objects (gray). Percent of signal change (PSC) from baseline was normalized to the maximum PSC for each ROI. Error bars represent standard error. Sigmoidal functions (Naka-Rushton (86)) fit to mean time courses are shown for all areas and conditions with significant fits as solid lines. (**D**) Steepness of the response (as taken from the Naka-Rushton function fit shown on Fig. 3C) for the different face preferential ROIs and an object preferential ROI in which the sigmoid function was fit successfully. Error bars depict 95% confidence intervals obtained by bootstrap. Asterisks represent significant differences (assessed by non-overlapping CIs) between familiar, unfamiliar faces and objects within individual ROIs: * $p<0.05$, ** $p<0.001$, all other comparisons are not significant. (**E**) Response steepness for the four grouped ROIs was analyzed with a 3 (stimulus type: familiar/unfamiliar faces/objects) x 4 (ROI: core, prefrontal and temporal extended, object) ANOVA. Error bars represent standard error. Response steepness depended on the interaction of stimulus type and ROI ($F(6,156) = 4.95$, $p < 0.0001$). Post hoc tests are shown with asterisks: * $p<0.05$, ** $p<0.01$, *** $p<0.001$, corrected using Tukey's honest significance difference. Comparisons 'ns' did not reach significance.*


A
Personally familiar
Unfamiliar
Objects
t = 0
cpi = 7
Increasing HSF
t = 32 s
cpi = 36.64
B
Face processing
Familiar face recognition
Brain activity
Time
Time
C
Relative % signal change
PL
ML
MF
MD
AL
AF
AD
AM
PO
PA
PV
Obj
TP
PR
Time (s)
D
Response steepness (a.u.)
PL ML MF MD AL AF AD AM PO PA PV PR TP Obj
E
Steepness (a.u.)
ns
Core system
Prefrontal
New anterior temporal
Object selective
Extended systems

The implications of the present results for our understanding of familiar face processing will be discussed in depth in Chapter IV. In brief, using whole-brain functional magnetic resonance imaging, we found that personally familiar faces engage the macaque face-processing network more than unfamiliar faces. Additionally, familiar faces recruited two previously unknown face areas at anatomically conserved locations within the perirhinal cortex and the temporal pole. These two areas, but not the core face-processing network, responded to familiar faces emerging from a blur with a characteristic nonlinear surge, akin to the abruptness of familiar face recognition. In contrast, responses to unfamiliar faces and objects remained linear. Thus, we found two temporal lobe areas that extend the core face-processing network into a familiar face-recognition system.

Chapter III.

Neural coding of familiar faces and individuals in the new temporal pole face area

As a first step towards advancing our understanding of the neuronal processing of familiar faces, our recent fMRI experiments in monkeys have provided evidence for two brain regions selective for familiar face-related content (see chapter II, (*87*)), one located in the perirhinal cortex (face area PR) and the other in the temporal pole (face area TP). Furthermore, a categorical and specific distinction between familiar and unfamiliar faces emerged in these areas, conforming to a pattern predicted for the recognition of familiar individuals. fMRI measures average blood flow within voxels that contain hundreds of thousands of neurons, and cannot directly assess the selectivity of single cells. Thus, the neurophysiological underpinnings of TP and PR remain unexplored. Likewise, it is not clear whether neuronal strategies for generating selective representations of familiar and unfamiliar faces might differ between these novel areas and other face areas.

Even though TP and PR were the only face areas showing a categorical distinction between familiar and unfamiliar faces, personal familiarity also modulated the activity of core face areas. These areas, even if purely perceptual, could be reflecting familiarity if they were tuned to physical parameters of familiar faces (*88-90*). If this was the case, unfamiliar faces that look like familiar faces should elicit a similar response in these areas. In this way, while TP and PR signal familiarity of faces in an all or none manner, the core face areas might just signal perceptual similarity to familiar faces.

Beyond familiarity with faces, neurons in TP and PR could be engaged in the act of individual recognition itself: cells in TP and/or PR might integrate information about familiar faces with other individual characteristics such as bodies and voices, to achieve a unique representation from each individual. Alternatively, face, body and voice selective cells could be intermixed, achieving individual recognition through specific connections. In any case, this would imply a transition of TP and PR from face areas to 'multimodal hubs' where individual recognition would take place (*1*).

Whether these newly discovered face areas are also sensitive to bodies or voices is unknown. Face area TP is an ideal candidate to investigate these questions. The monkey temporal pole is sensitive to species-specifc calls (*91*) and has been proposed as a multisensory convergence site (*92*). Moreover, humans with impaired multimodal recognition of familiar faces and voices usually have brain bilateral damage in this region (*50*). Given that we have functionally localized a region in the temporal pole that is selective to familiar faces, we are in a unique position to test if there are signs of multimodal representations for individuals in TP.

We therefore decided to perform electrophysiological recordings to test the hypothesis that TP supports the recognition of familiar individuals. To do so, we asked the following questions:

(1) Does TP contain face-selective cells? If so, are TP cells selective to familiar faces and does their selectivity depend on the physical parameters of faces?

(2) Do TP cells show hallmarks typical of familiar face recognition?

(3) Does TP contain body-selective cells and/or whole-individual selective cells?

(4) Does TP contain voice cells and/or multisensory cells?

Finally, we aimed at comparing TP neurons' activity to the activity of neurons recorded in face area AM, an area that has been hypothesized to develop a view-invariant internal model of faces (*93*), with the goal of understanding how cells that reside in a purely perceptual face-area compare with TP cells.

Single cell properties and population readout in TP and AM

Our fMRI study established areas TP and PR as sites specialized for familiar face processing and also put personal familiarity as a fundamental dimension to study in the field of face recognition (see Chapter II, Figure 6). Given that the fMRI signal does not allow for a direct assessment of neuronal activity, the neurophysiological properties of TP remain unexplored. A main goal of our electrophysiological experiments was thus to understand the selectivity of single neurons within this specific region of the temporal lobe. Considering that personal familiarity modulated the activity of the entire face processing system (see Chapter II, Figures 3-6), and that some AM cells represent unfamiliar face identities in almost a fully view-invariant way (*18*), a second key goal was to compare neuronal strategies subserving differences in familiar and unfamiliar face representations in TP's and AM's neuronal activity.

To explore face and familiarity selectivity of TP, and compare it to AM, we first recorded neural responses to a 95-image set consisting of 7 categories (13 images per category) and 4 gray background pictures. The categories were: personally familiar and unfamiliar macaque faces, personally familiar and unfamiliar human faces, personally familiar and unfamiliar objects, and unfamiliar bodies.

Electrophysiological recording sites were localized using the stereotaxic coordinates of the fMRI maps, targeting areas TP and AM in the right hemisphere of monkeys M1 and M2 (figure 7A). For monkey M1, the targeting of areas TP and AM was performed but only a few neurons in AM have been sampled so far. We tested for face and familiarity selectivity in each cell we encountered. Across the population of recorded cells, 62 of 73 (85%) TP neurons (all in M2) and 58 of 66 (88%) AM neurons (11 in M1, 55 in

M2) gave significant responses to at least one of the images and were therefore classified as visually responsive.

Figure 7. fMRI-guided electrophysiology of face areas TP and AM

*(**A**) Coronal planes showing recoding locations targeting TP (left) and AM (right) in monkeys M1 and M2. Electrodes targeting these areas were lowered through recording grids into the brain. The chambers and recording grids, filled with gadolinium, are also visible in the images. Green lines indicate the projected recording trajectories. The functional overlay shows the regions that responded significantly more to Faces than to Objects in the fMRI experiment (see Chapter II, Figure 3), color coded for negative common logarithm of p value (q<0.05, FDR corrected). The stereotaxic coordinates of targeted areas are provided in Table 2 (Chapter II). (**B**) PSTH of two example cells in TP (left) and AM (right) in response to faces and objects that were either personally familiar or unfamiliar, to unfamiliar bodies and to a gray background. Each row shows the PSTH for one image that was presented for 200 ms, followed by a 500 ms of gray background. The color coding indicates the mean firing rate of 5-10 repetitions of each image in Hz. (**C**) Mean spike density functions for each of the four cells showed in (B) averaged across images for each of the stimulus categories, following the color coding indicated in (A).*

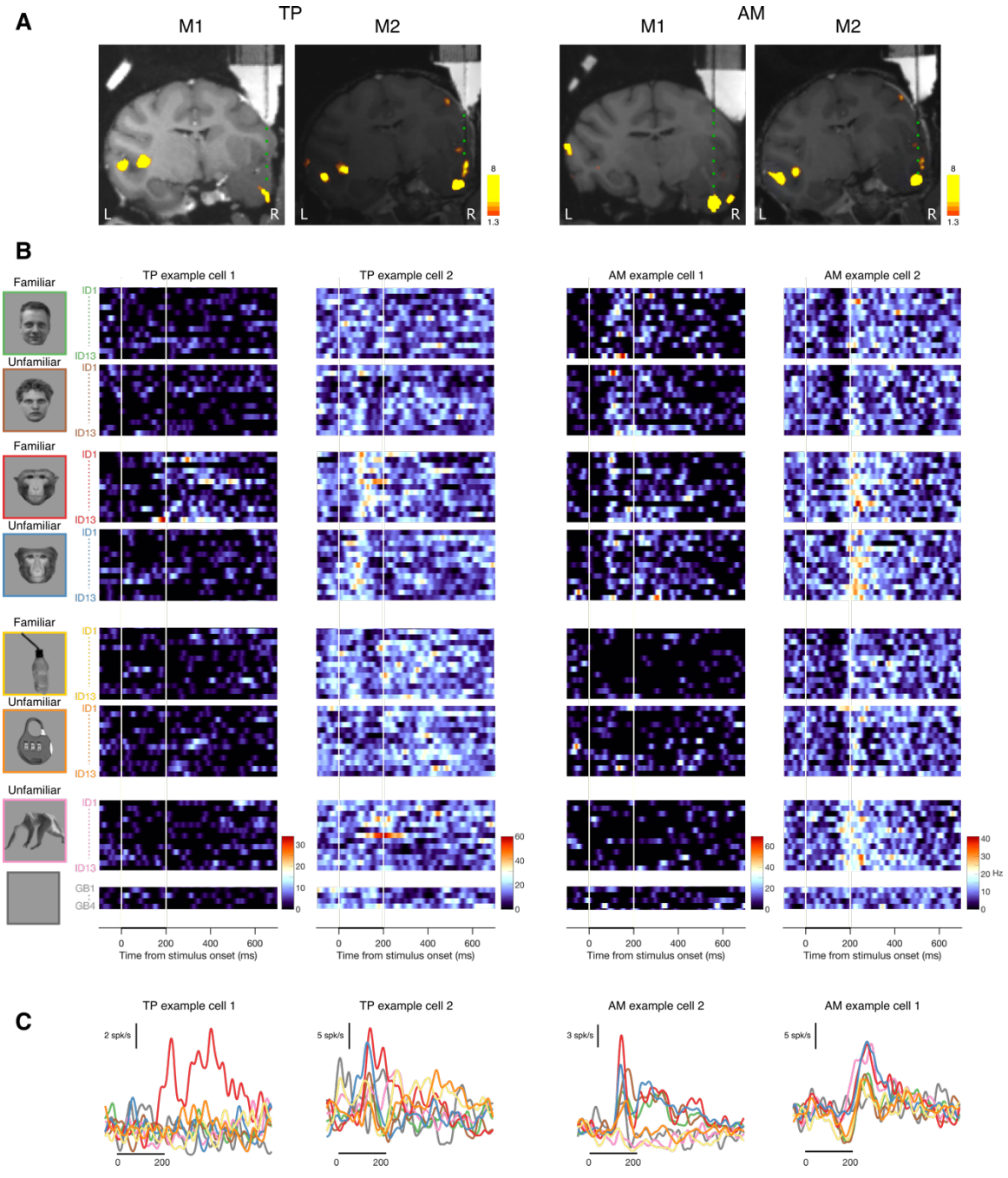

A
TP
AM
M1
M2
M1
M2
L
R
8
1.3
B
TP example cell 1
TP example cell 2
AM example cell 1
AM example cell 2
Familiar
Unfamiliar
Familiar
Unfamiliar
Familiar
Unfamiliar
Unfamiliar
ID1
ID13
GB1
GB4
30
20
10
0
60
40
20
0
60
40
20
0
40
30
20 Hz
10
0
0
200
400
600
Time from stimulus onset (ms)
C
TP example cell 1
TP example cell 2
AM example cell 2
AM example cell 1
2 spk/s
5 spk/s
3 spk/s
5 spk/s
0
200

Figure 7B shows the responses of two typical cells in TP to the 95-image stimulus set. Cell 1 responded strongly to multiple monkey faces that were familiar and was thus selective for the familiar monkey face category. Cell 2 responded both to faces of familiar and unfamiliar monkeys, as well as to monkey bodies, but not to human faces or objects. This was evident for single images (Figure 7B) as well as for the category averaged responses (figure 7C). In AM, most cells responded to faces of human and monkeys, such as the example cell 1 shown in Figure 7B. Other AM cells were selective for familiar and unfamiliar monkey faces and bodies, like example cell 4, which was most evident when averaging across images for each category (Figure 7C).

These selectivity profiles were typical for the entire population of cells recorded. Figures 8A and 9A show the response profiles of all cells recorded in TP and AM, respectively, to the 95-image stimulus set. In TP 96% of visually responsive cells were face selective, with 90% selectively enhanced by faces (average face response ≥ 2 * average non-face response) and 6% selectively suppressed by faces [computed Face Selectivity Index (FSI) with all face categories, Figure 8D]. When averaging across stimuli for the same species category, 60% of visually responsive cells were modulated by the familiarity of monkey faces (Figure 8E) and 30% by the familiarity of human faces (Figure 8F) [Familiarity Index, see Experimental procedures]. For each face-selective cell, we determined the stimulus that elicited the maximum response (i.e. the preferred face). A significant proportion of face-selective units (80%) preferred a familiar monkey face (Figure 8A and 8G), over other face stimuli (unfamiliar monkey faces, familiar or unfamiliar human faces). Thus, neurons in TP showed a remarkable specificity for face and familiarity processing.

In AM, 88% of visually responsive cells were face selective, with 60% selectively enhanced by faces and 28% selectively suppressed by faces (Figure 9D). The selectivity patterns in AM and TP differed: in TP, familiarity with monkey faces modulated responses in 60% of visually responsive cells, while in AM only 26% showed this same modulation (Figure 9E), with a similarly low fraction modulated by human face familiarity (23% Figure 9F). This was also reflected in the distribution of cells according to their preferred face: a smaller proportion of face-selective AM cells (37%) preferred a familiar monkey face compared to TP (80%), and the distribution of preferred faces in AM was even between the four face categories (Figure 9A and 9G)). Cells in AM showed face-selectivity as previously reported (*17*, *18*), but were not modulated by familiarity like cells in TP.

In TP, familiar monkey face selectivity was not only apparent in single-cells' spiking response, but also in the population (averaged over all cells, Figure 8B) and LFPs (Figure 8C). familiar monkey faces elicited a higher spiking response than all other categories from early in the stimulus presentation until beyond the stimulus offset. A significant preference for familiar monkey faces emerged on average 160 ms after stimulus onset (figure 8B, shaded area, $p<0.01$) in the spiking response and at 84 ms in the LFPs (figure 8C, shaded area, $p<0.05$). In AM, despite the fact familiar monkey faces also induced the highest spike rates during most of the stimulus presentation, this difference failed to reach statistical significance at any point, coming closest 190 ms after onset ($p=0.06$, Figure 9B). In AM, LFPs showed no significant modulation by familiarity at any time point (Figure 9C). Consistent with the observation at single-cell level, TP's neural population showed a remarkable specificity for face and familiarity processing that emerged early on and persisted beyond stimulus presentation.

\

Figure 8. Face selectivity of neural populations in TP.

*(**A**). Population response matrices to the Face and Familiarity Selectivity image set for all recorded cells in TP (n=73). Responses are sorted from top to bottom by the face selectivity indices (FSI calculated with all face stimuli, see Experimental procedures), and from left to right by stimulus category. (**B**) Average normalized TP population PSTHs for all categories. The color shading indicates the standard error of the mean (SEM) for each category. The gray shaded area indicates time intervals during which the population response to familiar monkey faces is significantly greater than to unfamiliar monkey faces (estimated using a bootstrap procedure, n=1000 iterations, ** p<0.01). (**C**) Average local-field potential (LFP) responses over all TP sites (n=34, mean and SEM) for each stimulus category, following the color coding indicated in (A). Shaded area indicates the time points with significant differences between familiar and unfamiliar monkey faces (estimated using a bootstrap procedure, n=1000 iterations, * p<0.05). (**D**) Face selectivity of neural population responses in TP, showing the distributions of FSIs for visually responsive cells, calculated with all face stimuli. (**E**) Monkey familiarity index for the visually responsive neural population in TP. (**F**) Human familiarity index for the visually responsive neural population in TP. (**G**) Category of the stimulus that elicited the highest response for cells in TP, following the color coding indicated in (A). (**H**) Population similarity matrix in TP. A 95 by 95 matrix of correlation coefficients was computed between responses of all cells to the 95 stimuli. (**I**) Multidimensional scaling plot of responses to the 95 stimuli within the first two dimensions of the MDS space, following the color coding indicated in (A). Lines were just plotted to mark the boundaries of each category.*

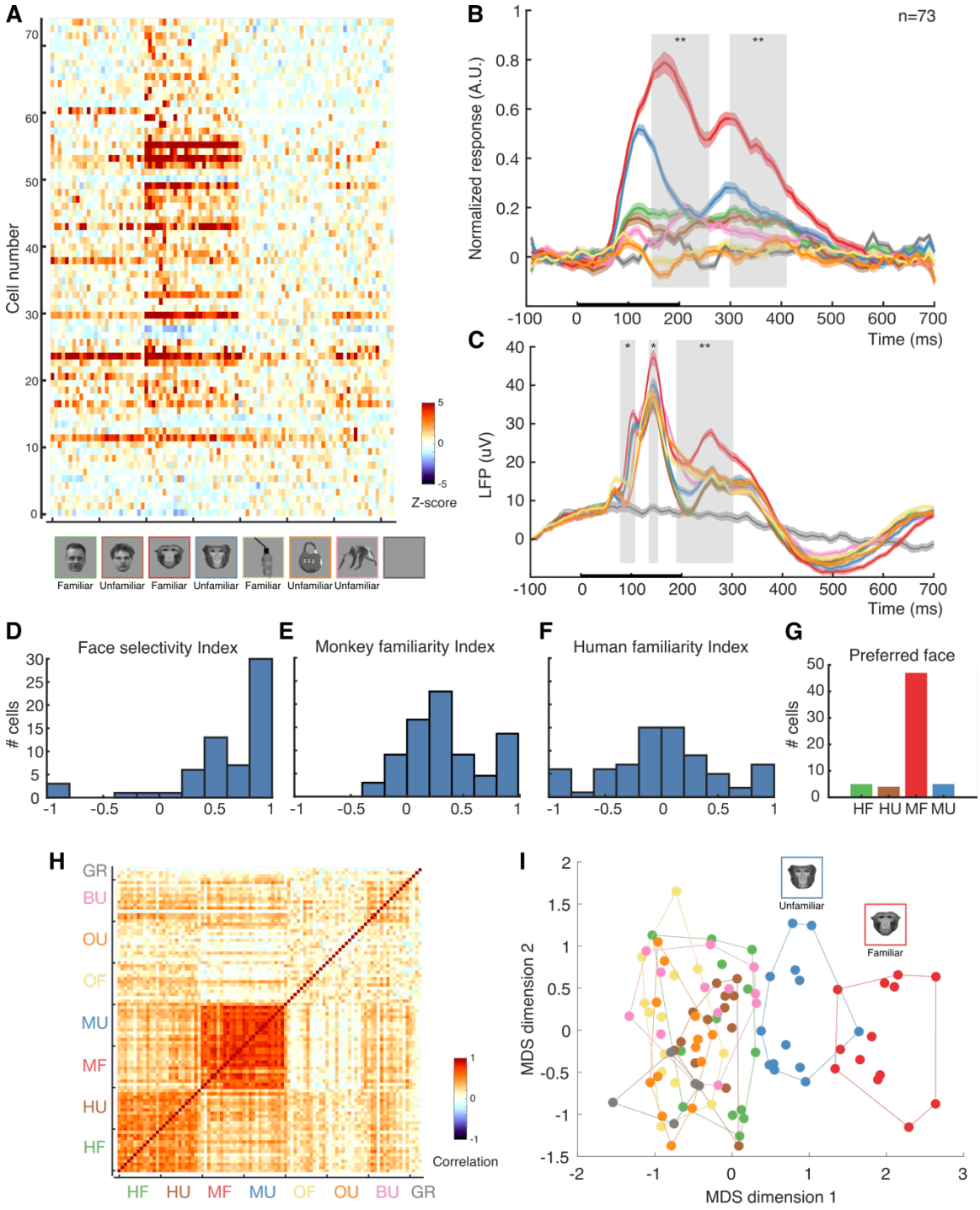
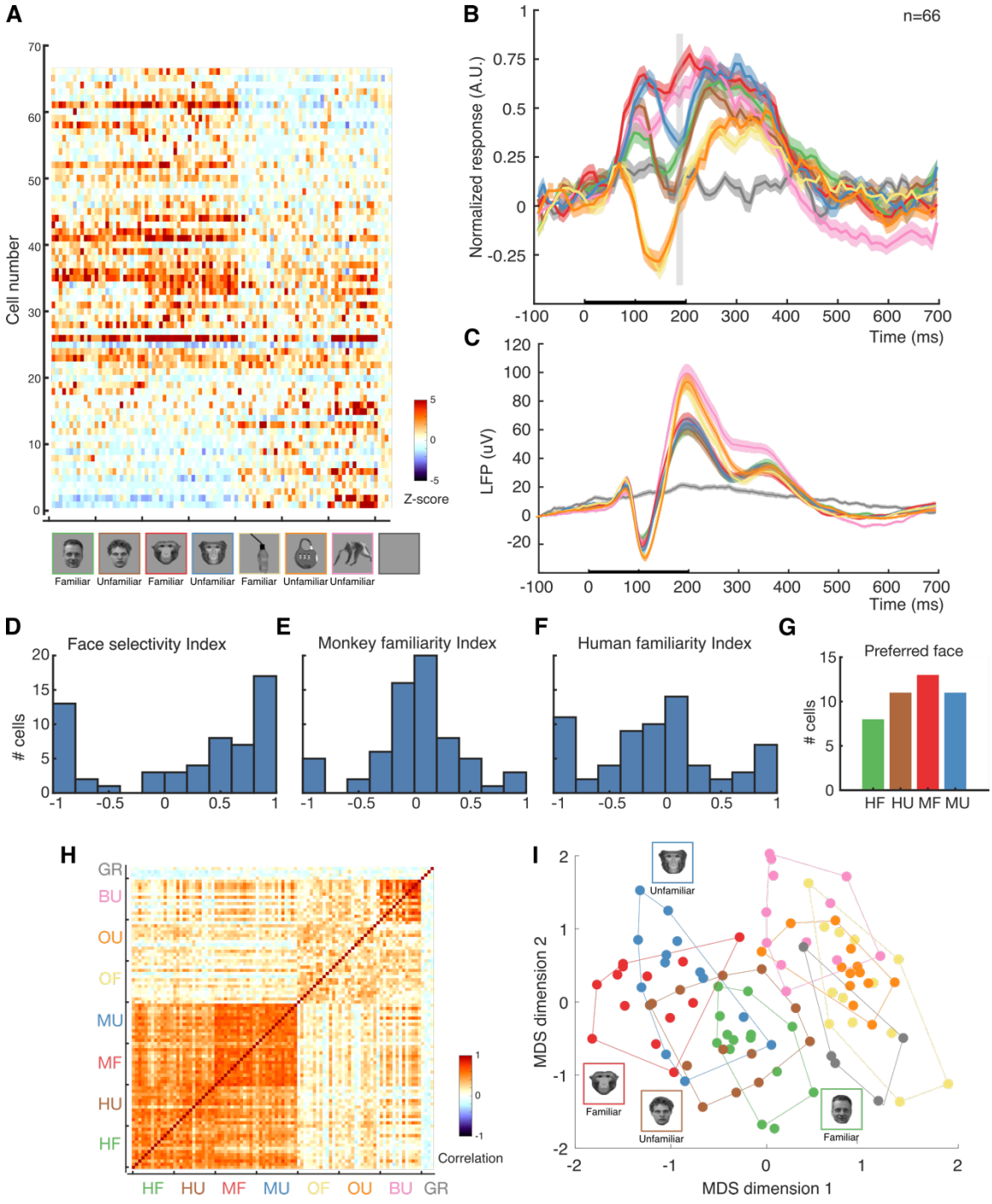

A
Cell number
Z-score
Familiar
Unfamiliar
Familiar
Unfamiliar
Familiar
Unfamiliar
Unfamiliar
B
n=73
Normalized response (A.U.)
Time (ms)
C
LFP (uV)
Time (ms)
D
Face selectivity Index
cells
E
Monkey familiarity Index
F
Human familiarity Index
G
Preferred face
cells
HF HU MF MU
H
GR
BU
OU
OF
MU
MF
HU
HF
Correlation
HF HU MF MU OF OU BU GR
I
Unfamiliar
Familiar
MDS dimension 2
MDS dimension 1

Figure 9. Face selectivity of neural populations in AM.

*(**A**). Population response matrices to the Face and Familiarity Selectivity image set for all recorded cells in AM (n=66). Responses are sorted from top to bottom by their face selectivity indices (FSI calculated with all face stimuli, see Experimental procedures), and from left to right by stimulus category. (**B**) Average AM normalized population spiking response to all categories. The color shading indicates the standard error of the mean (SEM) for each category. The gray shaded area indicates time intervals in the mean spiking response during which the population preference for familiar monkey faces is marginally significant (estimated using a bootstrap procedure, n=1000 iterations, * p=0.06). (**C**) Average local-field potential (LFP) responses over all AM sites (n=34, mean and SEM) for each stimulus category, following the color coding indicated in (A). No significant differences were found at any time point between familiar and unfamiliar monkey faces (estimated using a bootstrap procedure, n=1000 iterations, p>0.1). (**D**) Face selectivity of neural population responses in AM, showing the distributions of FSIs for visually responsive cells, calculated with all face stimuli. (**E**) Monkey familiarity index for the visually responsive neural population in AM. (**F**) Human familiarity index for the visually responsive neural population in AM. (**G**) Category of the stimulus that elicited the highest response for cells in AM, following the color coding indicated in (A). (**H**) Population similarity matrix in AM. A 95 by 95 matrix of correlation coefficients was computed between responses of all cells to the 95 stimuli. (**I**) Multidimensional scaling plot of responses to the 95 stimuli within the first two dimensions of the MDS space, following the color coding indicated in (A). Lines were plotted to mark the boundaries of each category.*


A
Cell number
Z-score
Familiar
Unfamiliar
Familiar
Unfamiliar
Familiar
Unfamiliar
Unfamiliar
B
n=66
Normalized response (A.U.)
Time (ms)
C
LFP (uV)
Time (ms)
D
Face selectivity Index
cells
E
Monkey familiarity Index
F
Human familiarity Index
G
Preferred face
cells
HF HU MF MU
H
GR
BU
OU
OF
MU
MF
HU
HF
Correlation
HF HU MF MU OF OU BU GR
I
Unfamiliar
Familiar
Unfamiliar
Familiar
MDS dimension 2
MDS dimension 1

To determine which parameters are driving neuronal activity at the population level in both brain areas, we measured the similarity (correlation) between population responses to all images, that is, to all combinations of stimulus pairs to construct population response similarity matrices. The TP similarity matrix (Figure 8H) reflected a high similarity for responses to monkey faces, and a separate patch of high similarity for human faces. The fact that we did not find a separate patch for familiar monkey faces is an indication that different cells in TP might encode different familiar identities (see section below about Selectivity and Sparseness). In contrast, the AM similarity matrix is characterized by a bigger patch (figure 9H) reflecting a high-similarity between human and monkey face population response vectors. While in TP, neuronal activity was driven differently by monkey and human faces, AM showed a similar population response pattern for human and monkey faces.

To visually represent the population responses to all stimuli, we applied multidimensional scaling (MDS) to the two populations. MDS represent the population responses to the stimuli in a reduced number of dimensions, while preserving the Euclidean distances between points. Two-dimensional representations were derived (Figures 8F and 9F), allowing the proximity in the plot to reflect the similarity of the pattern of responses across the neural population. A consequence of the emergence of familiarity tuning in TP is that the neural population becomes good at differentiating familiar and unfamiliar monkey faces, and also at differentiating monkey faces from the rest of the stimuli (Figure 8I). In AM, face and non-face stimuli formed separate clusters based on species (Figure 9I). In TP, this clustering is broken, and responses to human faces are collapsed into the cluster of non-face stimuli.

Cells in the temporal pole face area were mostly face-selective and specifically responsive to familiar monkey faces. Because many cells were activated by a few familiar face stimuli, the responses of cells in TP tended towards a sparser code than responses of cells in AM. We thus decided to run a larger stimulus set to quantify this better (see below).

Selectivity and sparseness of face-selective cells

To explore the selectivity and sparseness of TP further, we probed face-selective cells with a larger stimulus set. Only cells for which at least 6 repetitions/stimulus surpassed our behavioral criteria were considered for the analysis (n=30 in TP, n=36 in AM, see Experimental procedures). We used the same categories of visual stimuli but we added pictures of 59 unfamiliar monkey faces, 17 familiar and unfamiliar human faces, 2 familiar objects, 10 unfamiliar objects, and 2 unfamiliar bodies, resulting in an expanded stimulus set of 205 images.

Figure 10A shows the responses of two typical cells in TP to the 205-image stimulus set. Both cells prefer monkey faces over all other stimuli. However, whereas TP example cell 3 responded strongly to different familiar monkey faces, and to a lesser degree to some unfamiliar monkey faces and bodies, TP example cell 4 was highly selective, firing to only a single familiar monkey's face. This strong selectivity was consistent across trials (Figure 10B). The responses of these cells in TP were representative of the population (Figure 10C). Only once cell with similar characteristics was found in AM, shown in Figure 11A (AM example cell 3): this cell responded strongly to a familiar human face and a familiar monkey face, and to a lesser degree to other faces (see consistency across trials in

Figure 10B). However, this was an exceptional cell among the recorded cells in AM, and most of them resembled example cell 4 (see Figure 11C for the population responses).

Figure 10. Selectivity and sparseness of neural populations in TP.

*(**A**). PSTH of example cells in TP in response to an extended version of the stimuli set, consisting of faces and objects that were either personally familiar or unfamiliar, to unfamiliar bodies and to a gray background. Each row shows the PSTH for one image that was presented for 200 ms, followed by a 500 ms of gray background. The color coding indicates the mean firing rate in Hz. (**B**) Raster plot showing the occurrences of spikes in each trial for the first 12 preferred images of the example cell 4. (**C**) Population response matrices to the extended 205-image set for all face-selective cells in TP. (**D**) Sparseness of tuning of face-selective cells in TP, quantified by the distributions of identity sparseness indices computed from responses of cells to the 145 identities of faces (human and monkey faces) in the stimuli set. A lower sparseness index indicates sharper identity tuning. (**E**) Distribution of response selectivity across face-selective cells in TP, computed as the percentage of stimuli eliciting responses larger than half of the maximum (>HM) response for each cell.*

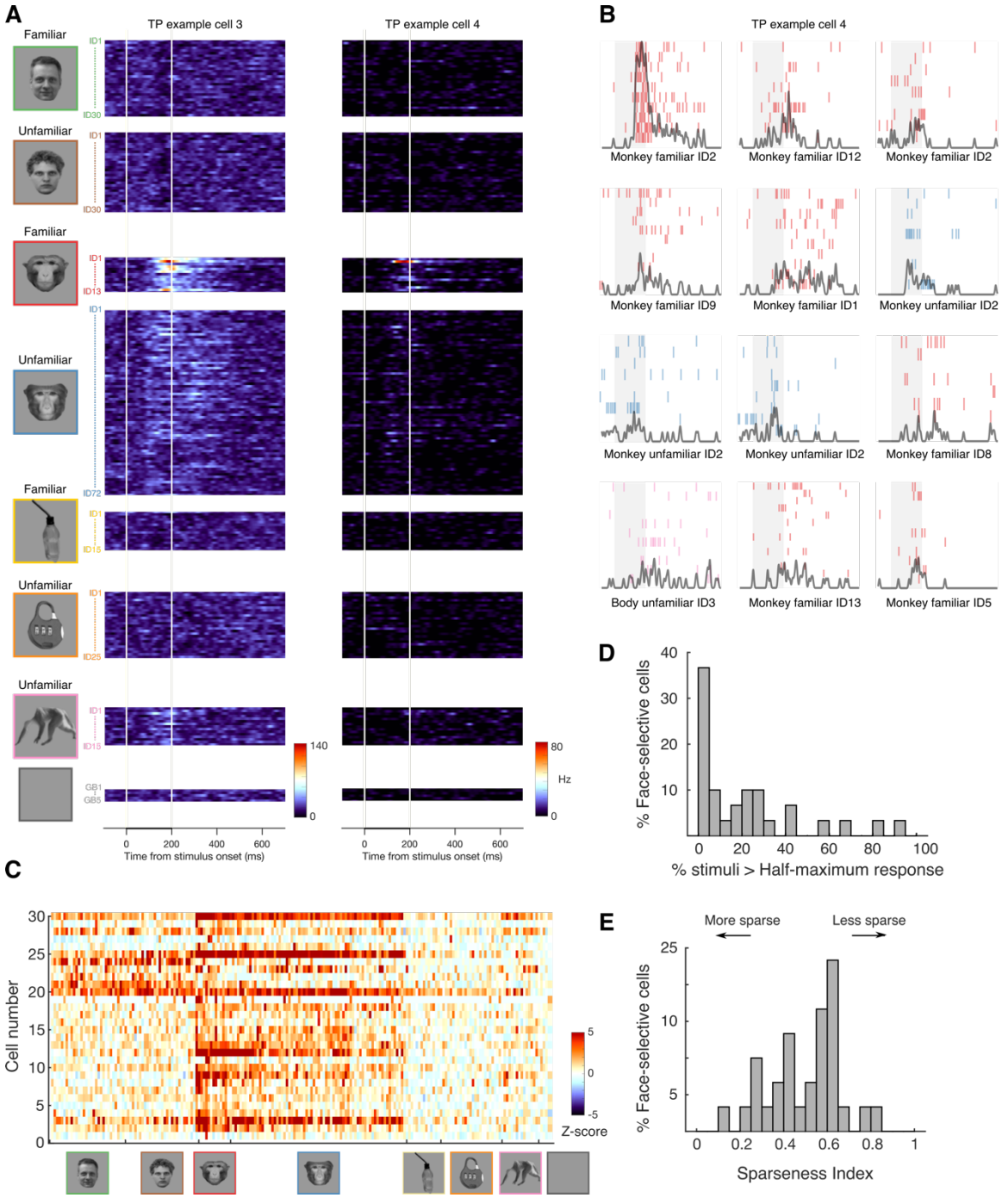
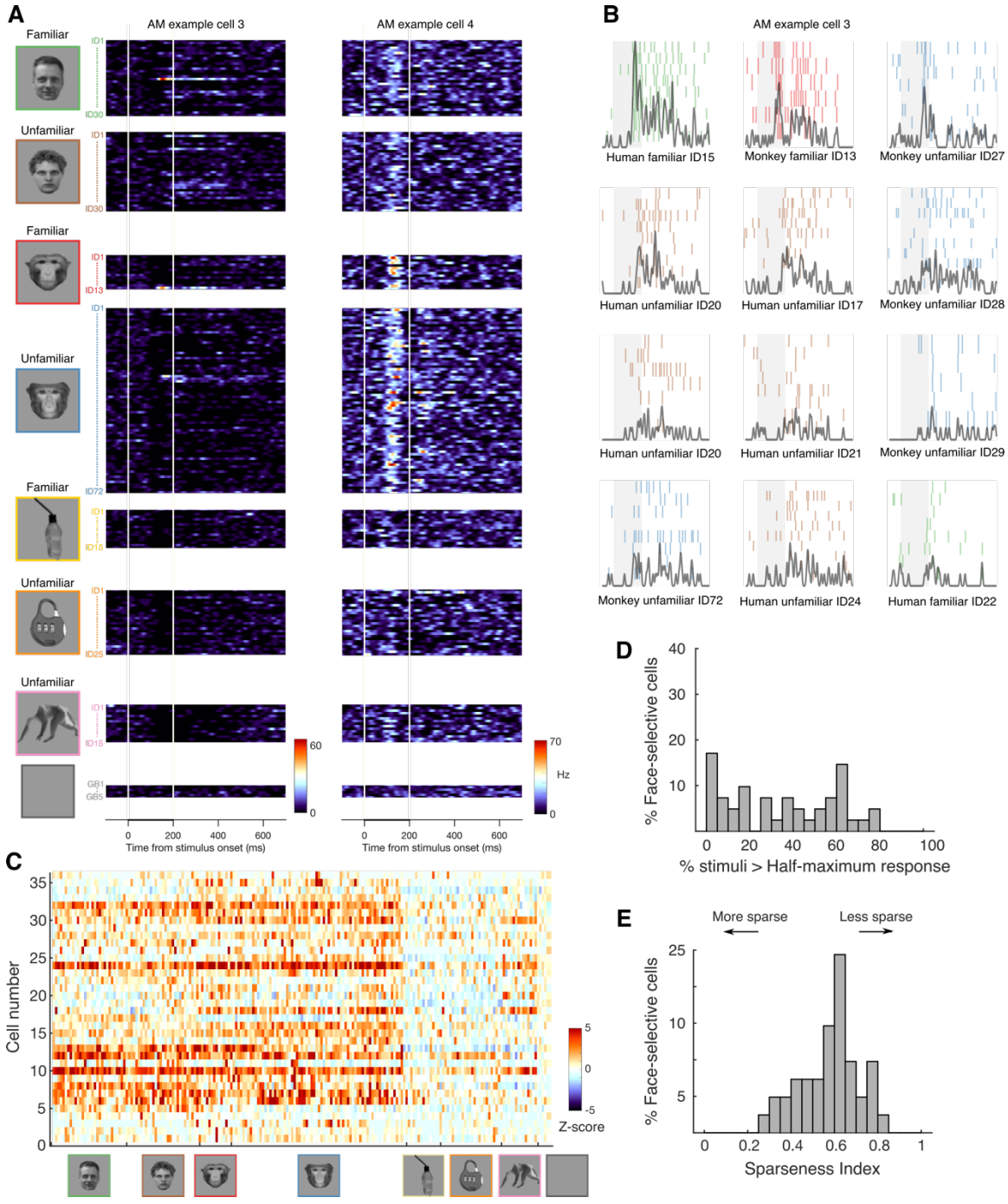

A
TP example cell 3
TP example cell 4
Familiar
Unfamiliar
Familiar
Unfamiliar
Familiar
Unfamiliar
Unfamiliar
ID1
ID30
ID1
ID30
ID1
ID13
ID1
ID72
ID1
ID15
ID1
ID25
ID1
ID18
GB1
GB5
140
0
80
Hz
0
0
200
400
600
Time from stimulus onset (ms)
B
TP example cell 4
Monkey familiar ID2
Monkey familiar ID12
Monkey familiar ID2
Monkey familiar ID9
Monkey familiar ID1
Monkey unfamiliar ID2
Monkey unfamiliar ID2
Monkey unfamiliar ID2
Monkey familiar ID8
Body unfamiliar ID3
Monkey familiar ID13
Monkey familiar ID5
C
Cell number
Z-score
D
% Face-selective cells
% stimuli > Half-maximum response
E
More sparse
Less sparse
% Face-selective cells
Sparseness Index

Figure 11. Selectivity and sparseness of neural populations in AM.

*(**A**). PSTH of example cells in AM in response to an extended version of the stimuli set, consisting of faces and objects that were either personally familiar or unfamiliar, to unfamiliar bodies and to a gray background. Each row shows the PSTH for one image that was presented for 200 ms, followed by a 500 ms of gray background. The color coding indicates the mean firing rate in Hz. (**B**) Raster plot showing the occurrences of spikes in each trial for the first 12 preferred images of the example cell 4. (**C**) Population response matrices to the extended 205-image set for all face-selective cells in AM. (**D**) Sparseness of tuning of face-selective cells in AM, quantified by the distributions of identity sparseness indices computed from responses of cells to the 145 identities of faces (human and monkey faces) in the stimuli set. A lower sparseness index indicates sharper identity tuning. (**E**) Distribution of response selectivity across face-selective cells in AM, computed as the percentage of stimuli eliciting responses larger than half of the maximum (>HM) response for each cell.*


A
AM example cell 3
AM example cell 4
Familiar
Unfamiliar
Familiar
Unfamiliar
Familiar
Unfamiliar
Unfamiliar
ID1
ID30
ID13
ID72
ID18
ID25
ID15
GB1
GB5
60
0
70
Hz
0
200
400
600
Time from stimulus onset (ms)
B
AM example cell 3
Human familiar ID15
Monkey familiar ID13
Monkey unfamiliar ID27
Human unfamiliar ID20
Human unfamiliar ID17
Monkey unfamiliar ID28
Human unfamiliar ID20
Human unfamiliar ID21
Monkey unfamiliar ID29
Monkey unfamiliar ID72
Human unfamiliar ID24
Human familiar ID22
C
Cell number
35
30
25
20
15
10
5
0
5
0
-5
Z-score
D
% Face-selective cells
40
30
20
10
0
20
40
60
80
100
% stimuli > Half-maximum response
E
More sparse
Less sparse
% Face-selective cells
25
10
5
0
0.2
0.4
0.6
0.8
1
Sparseness Index

We quantified the selectivity of face-selective cells to compare the encoding properties between TP and AM. In TP, face-selective cells responded to an average of 17% of all the visual stimuli (that is, 36/205 of the stimuli elicited responses greater than the half-maximum response), and 43% of the cells responded to an average of less than 5% of the visual stimuli (Figure 10D). AM cells responded to an average of 28% of the visual stimuli (that is, 57/205 of the visual stimuli elicited responses greater than the half-maximum response), and 24% of the cells responded to an average of less than 5% of the visual stimuli (Figure 11D). The differences in selectivity between AM and TP were significant (estimated by a bootstrap procedure, p=0.02, 1000 iterations), indicating that the population of TP cells was more selective than the population of AM cells.

To compare the sparseness of the neural encoding of faces, we computed a sparseness index (*34*) with the responses to all face stimuli. This index ranges from 0 (sparse coding) to 1 (dense coding). The mean sparseness index for TP cells was 0.47 whereas for AM cells was slightly higher (0.61, i.e. less sparse) (Figure 10E and Figure 11E), however, there was not significant difference between TP and AM (estimated by a bootstrap procedure, p=0.1, 1000 iterations),

Taken together, these comparisons suggest that (1) TP cells are more selective than AM-cells and (2) that TP and AM cells rely on an intermediate coding strategy, with TP cells having a tendency for sparser codes.

Critical parameters driving preference for familiar faces in TP

Both single-cell and neural population responses in TP showed a strong preference for familiar faces (see Figure 8 and Figure 10). Considering that experience shapes cortical representations (*88-90*) and that our fMRI results point towards a modulation of the entire face-processing system by familiarity, our findings raised the following question: what are the critical parameters driving the preference? Neurons in TP, like neurons in other face areas, could be sensitive to physical aspects of familiar faces, which over real-life exposure shaped the whole system.

To study the impact of physical characteristics in the responses of TP face-selective cells, and compare it with the impact they have in AM, the physical attributes of monkey faces were measured and related to the neuronal responses. We ran this analysis on cells for which we have both tested the extended image-set and whose preferred face was a monkey face (n=27 in TP, n=17 in AM). We labelled the most effective face stimulus for each cell as the preferred face. To determine whether individual cells were tuned to the physical aspects of faces we asked: how does the response of the most physically similar face to the preferred face compare to the response to the preferred face?

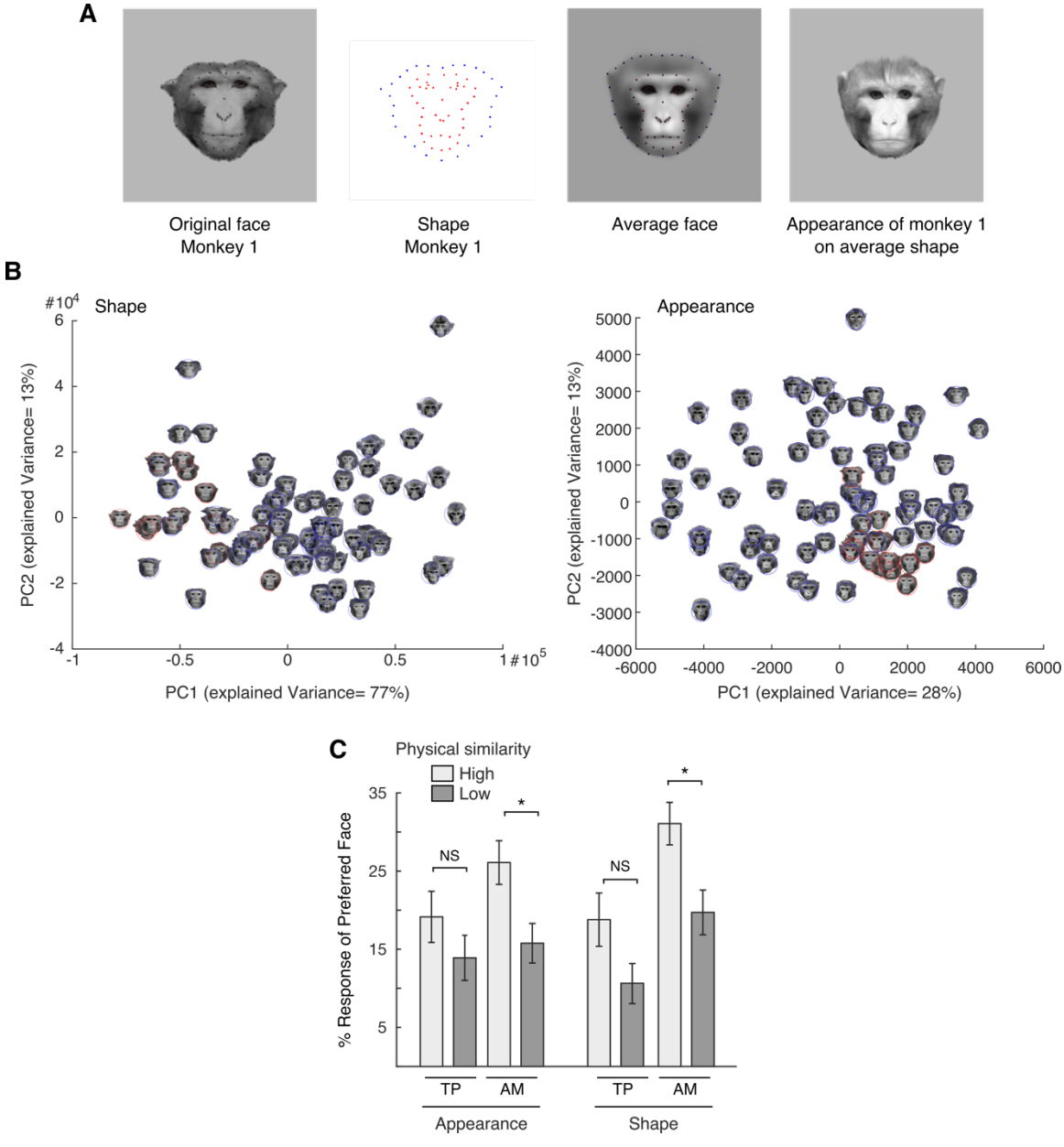
To measure physical similarity between faces, we followed a computational approach that separates physical aspects of faces in "shape" and "appearance" (*94*). Briefly, a set of anatomical landmarks were labeled by hand for each of the 85 monkey frontal faces in the extended image set (13 familiar, 72 unfamiliar) (Figure 12A, left). The positions of these points carry information about the shape of the face and the position of internal features (Figure 12A, middle left). The landmarks were then smoothly morphed to the average shape of all landmarks (Figure 12A, middle right), with the resulting image (Figure

12A, right) carrying "shape-free" appearance information. For these "shape-free" images one cannot simply use shape to discriminate them, and the physical parameters that are left include texture, reflectance, and resolution. We will use the term "appearance" as a shorthand for all these parameters. Separation of face shape and appearance is usually performed as a computational convenience, because it allows faces to be normalized before statistical treatments such as principal component analysis (PCA). In the hope of reducing the dimensionality, we derived the first two principal components for shape and appearance (Figure 12C, left and right respectively). While these two principal components of shape dimensions explained 90% of the variability, the first two principal components of appearance dimensions only accounted for 41% of the variability, and required 34 additional components to account for 90% variability. Thus, the full Euclidean distance was used as a measure of dissimilarity in physical parameters between faces.

In TP, faces that were physically similar to the preferred face failed to elicit a similar response: on average, the response was only 20% of the response to the preferred face (Figure 12C). This was not different for faces that bore low physical resemblance to the preferred face (estimated using a bootstrap procedure, n=1000 iterations, p=0.28 and p=0.08 for appearance and shape, respectively). In AM, this difference was marginally significant (p=0.05 for both appearance and shape parameters), and faces that were similar to the preferred face elicited a slightly higher, but not significant, response than in TP. Thus, there was a tendency for physical aspects of faces to impact activity of AM cells, but not in TP.

Figure 12. Shape and appearance features.

*(**A**) 70 landmark points were labeled on the 85 monkey face images (left). The positions of these landmarks carry information about the shape for each facial image (middle left). The landmarks were smoothly morphed to match the average landmark positions of the 85 faces (middle right), generating an image containing shape-free appearance information about each monkey face (right). (**B**) PCA was performed to extract the feature dimensions that account for the largest variability in the database. The first two components for shape and appearance are shown, familiar faces are plotted with a red circle surrounding them, and a blue circle was plotted around the unfamiliar faces. (**C**). Average and SEM response of the 5 faces with highest physical similarity in shape and appearance to the preferred face (light gray) and of the 5 faces that were more different to the preferred face (i.e. low physical similarity, dark gray). Response was calculated separated for each cell, expressed as a percentage of the response to the preferred face, and averaged across cells.*


A
Original face
Monkey 1
Shape
Monkey 1
Average face
Appearance of monkey 1
on average shape
B
Shape
#10^4
PC2 (explained Variance= 13%)
PC1 (explained Variance= 77%)
#10^5
Appearance
PC2 (explained Variance= 13%)
PC1 (explained Variance= 28%)
C
Physical similarity
High
Low
% Response of Preferred Face
NS
*
NS
*
TP
AM
TP
AM
Appearance
Shape

Hallmarks of familiar face recognition in TP

Most cells in TP were modulated by familiarity (see Figures 8A, 8C, 8D, 10C), and some of them were ultra-sparse (see example cell 4 in Figure 10A), responding only to a few personally familiar faces. However, unfamiliar faces that were similar in shape or appearance failed to elicit a similar level of response (Figure 12C). Together, these results suggest that TP cells implement a critical stage in the process of familiar face recognition beyond processing faces as visual stimuli. With two aims in mind, we tested whether we could detect hallmarks for familiar face recognition in TP cells: (1) to reproduce the 'aha' neural recognition effect we found with fMRI (Figure 6) at the single-cell level and, (2) to test several signatures of familiar face representations that were established by human behavioral studies, such as their resilience to distortions (*95-97*), the prevalence of internal features (*98-101*), and their likely viewpoint invariance nature (*1*).

Recognizing a familiar face: capturing the 'aha!' neural moment

As explained in Chapter II, poor image quality, induced e.g. by Gaussian blurring, disrupts recognition of unfamiliar faces, but not familiar faces (*82*, *102*, *103*). In our fMRI experiment, initially highly blurred and unrecognizable stimuli slowly incorporated, over the course of half a minute, increasing amounts of high spatial frequency (HSF) information (Figure 6A). This mode of stimulation allowed us to isolate the familiarity-related response, the 'aha!' neural recognition moment, from transient visual response(s) at a time scale reasonable for blood oxygenation responses (Figure 6C). TP showed a nonlinear response as blurred familiar faces gradually became visible, rapidly becoming active when faces of familiar monkeys became recognizable. The activation boost by

recognition did not seem to be fed back into the core system: in AM, as in other face areas, activity ramped up linearly as familiar and unfamiliar faces became unblurred.

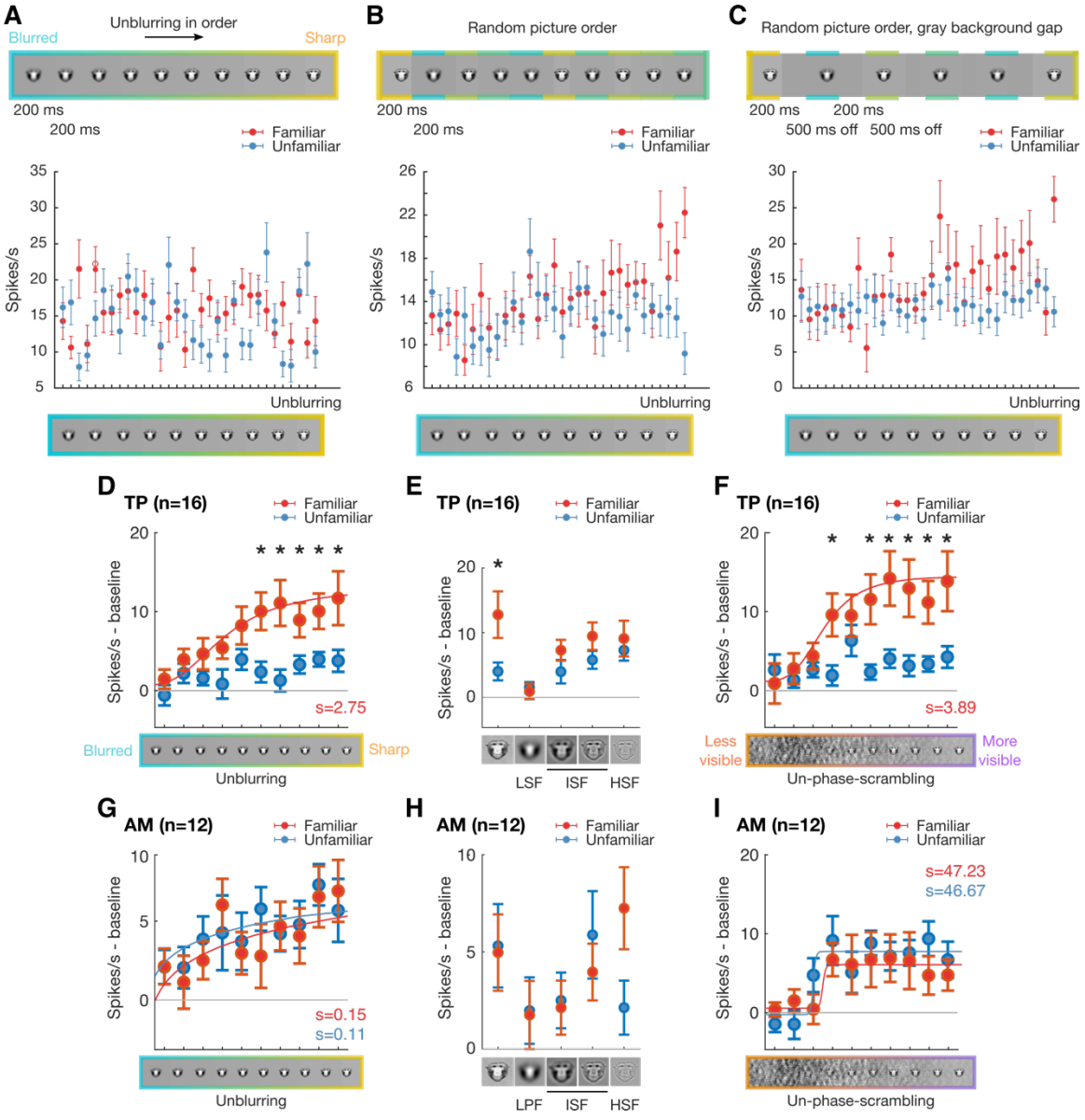
We aimed at replicating our fMRI result with electrophysiology, however, we encountered difficulties related to the different timescale resolution and to visual adaptation effects (*104*). First, we had to adjust the experimental paradigm. In order to capture the recognition moment with electrophysiology, we had to first adjust the time scale of the unblurring. Single cells in the temporal cortex show 'visual adaptation' effects: when the presentation of a stimulus is repeated several times, neurons decrease the strength of their response (*104*). When slowly unblurring a face, it is reasonable to expect an adaptation effect induced by showing very similar images one after the other. Considering the sharp tuning of some cells in TP, we selected the familiar and unfamiliar monkey faces that elicited the highest response for each neuron (i.e. the preferred face). Adaptation effects are stronger for stimuli to which cells are selective than for non-preferred stimuli (*105*). Thus, if the recognition of a familiar face also induces an activation boost at a single cell level, this effect could be masked by the adaptation effect, which would be stronger for the preferred face. This can be appreciated comparing Figures 13A, 13B and 13C. When unblurring faces in sequence without an interstimulus interval (ISI), there was no difference between the response to familiar and unfamiliar faces, and between different steps of the unblurring (Figure 13A). When switching to a random order presentation there was a sudden jump in the firing rate of the neuron, at the final steps of the unblurring (Figure 13B). Adaptation effects with a random order presentation are smaller, because pictures that follow each other are not as similar as when the unblurring occurs in order. Finally, when adding an ISI the sudden recognition effect became as clear as with the fMRI

paradigm (Figure 13C). The paradigm for electrophysiology was thus settled to a random picture presentation with an ISI that prevented a masking of the non-linear effect by visual adaptation.

A difference between the responses elicited by blurred familiar and unfamiliar faces was evident in both the single-cell (Figure 13C) and in the averaged population response for TP from intermediate blurring levels onwards (Figure 13D). AM cells failed to show this effect: the average population response to familiar and unfamiliar faces increased more or less linearly as faces became more distinguishable (Figure 13G). Reproducing our fMRI findings, TP showed an all-or-nothing response that evokes the sudden 'aha' moment we experience when we recognize a familiar face.

Figure 13. Neural signatures of familiar face recognition at a single-cell level in TP

*(**A**), (**B**), (**C**) Firing rate of an example cell recorded in TP with three different paradigms in which 32 images of a familiar (in red) and unfamiliar (in blue) face with different levels of blurring were shown. In (A) images were presented in order: the familiar and unfamiliar face were unblurred over the course of 6.4 seconds (200 ms on, no interstimulus interval - ISI). In (B) the same images were presented in a random order also every 200 ms and without an ISI. In (C) the same images were presented in a random order, every 200 ms, but with an ISI of 500 ms. (**D**) Average TP population spiking response to the unblurring paradigm shown in (C) with 10 steps of unblurring. Sigmoidal functions [Naka-Rushton (*86*)] fit to mean firing rate are shown for conditions with significant fits as solid lines. The steepness of the response estimated from this fit is shown for familiar ('s'). (**E**) Average TP population spiking response to 10 images with different levels of phase-unscrambling. Sigmoidal functions [Naka-Rushton (*86*)] fit to mean firing rate are shown for conditions with significant fits as solid lines. The steepness of the response estimated from this fit is shown ('s'). (**F**) Average TP population spiking response to the original images, and to four images with different spatial frequency content: low-spatial frequency (LSF), intermediate low and high frequencies (ISF), and high-spatial frequency (HSF) (see Experimental procedures). (**G**), (**H**), and (**I**) show AM population spiking response in the same way as in (D), (E), and (F) respectively. In (A), (B), and (C) error bars depict the standard error of the mean, while in (D)-(I) error bars depict the standard error of the mean (SEM). Significant differences between the response elicited by familiar and unfamiliar faces were estimated by a bootstrap procedure, 1000 iterations, * $p<0.05$.*


A
Unblurring in order
Blurred
Sharp
200 ms
200 ms
Familiar
Unfamiliar
Spikes/s
Unblurring
B
Random picture order
200 ms
200 ms
Familiar
Unfamiliar
Spikes/s
Unblurring
C
Random picture order, gray background gap
200 ms
200 ms
500 ms off
500 ms off
Familiar
Unfamiliar
Spikes/s
Unblurring
D
TP (n=16)
Familiar
Unfamiliar
Spikes/s - baseline
s=2.75
Blurred
Sharp
Unblurring
E
TP (n=16)
Familiar
Unfamiliar
Spikes/s - baseline
LSF
ISF
HSF
F
TP (n=16)
Familiar
Unfamiliar
Spikes/s - baseline
s=3.89
Less visible
More visible
Un-phase-scrambling
G
AM (n=12)
Familiar
Unfamiliar
Spikes/s - baseline
s=0.15
s=0.11
Unblurring
H
AM (n=12)
Familiar
Unfamiliar
Spikes/s - baseline
LPF
ISF
HSF
I
AM (n=12)
Familiar
Unfamiliar
Spikes/s - baseline
s=47.23
s=46.67
Un-phase-scrambling

The unblurring paradigm incorporates high frequency information to low-pass filtered images. The response profile in TP could thus be explained if TP cells were particularly sensitive to high-spatial frequencies. To study how cells in TP and AM respond to different spatial frequency bands, we created images of the preferred familiar and unfamiliar face that contained only low spatial frequency components (LSF, < 5 cycles/face), intermediate spatial frequencies (low and high ISF, 5-10 and 10-20 cycles/face respectively) and high spatial frequency components (HSF, >20 cycles/face) (see Experimental procedures for details). Even though there was a tendency for familiar faces in TP to elicit a higher response than unfamiliar faces in all but the extremely low-pass filtered images, this difference did not reach a significant level (Figure 13E). Interestingly, in AM high-pass filtered familiar faces elicited a higher response than unfamiliar faces, although this difference was also not significant (Figure 13H). Thus, while TP's non-linear effect was not caused by a special selectivity for high spatial frequencies, these spectrum components might modulate activity in AM.

Given the potential confounds of spatial frequency, we tried to capture the 'aha' neuronal moment with a different paradigm. Phase scrambling is typically used to retain low-level image properties, while making the content of the image less visible. In the same way as for the unblurring experiment, stimuli with different levels of phase scrambling were created for familiar and unfamiliar faces. The effect in TP was similar to the unblurring effect: for highly scrambled faces there was no significant response (Figure 13F), yet there was a sudden transition at intermediate levels of phase scrambling that remained unchanged for highly visible familiar faces. In AM, the phase-unscrambling effect was a sudden transition at intermediate levels of phase scrambling that remained

unchanged for highly visible faces (Figure 13I). Unlike cells in TP, this effect was similar for both familiar and unfamiliar faces.

We did not expect to see the highly non-linear effect in AM in the phase-scrambling experiment. Given that it was present for both familiar and unfamiliar faces, AM cells seem to have a threshold of visibility under which an image is not perceived as a face. The non-linearity in TP is different in the sense that is specific to familiar faces: cells respond only when sufficient evidence of a familiar face is available.

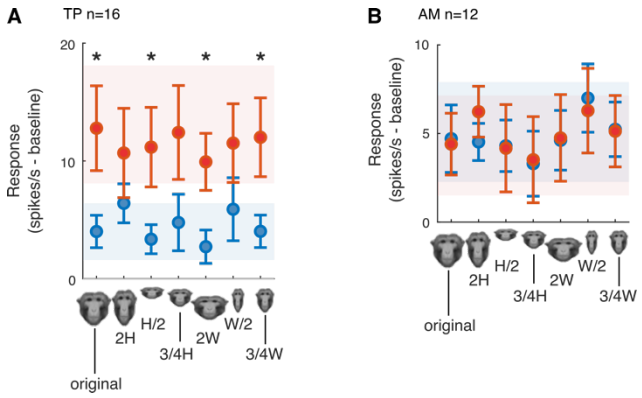
TP and AM cells tolerate distorted faces

Another hallmark of familiar face recognition is that it is remarkably robust under a range of deformations. One of the most powerful demonstrations shows that horizontal or vertical stretching images (making them too wide or too tall) has no effect on the recognition of familiar faces (*95-97*). To study this effect at the neuronal level, we tested the response of TP and AM cells to stretching and compressing familiar faces in the horizontal and vertical directions.

We found that the response of TP cells to familiar faces was unaffected by vertical and horizontal distortions of face images (Figure 14A). In AM, responses to both familiar and unfamiliar faces tolerated vertical and horizontal distortions (Figure 14B). Discounting these severe deformations thus occurs already in AM, and might even be a property of the perceptual system in earlier stages of computation.

Figure 14. TP and AM cells tolerate image distortions

*(**A**) Average TP population spiking response to the undistorted image, and to six distorted images of the preferred familiar and unfamiliar face of each cell. From left to right, responses to the following images are shown: original face image, stretched version of the original image to double its height, shortened version of the original image to half its height, shortened version of the original image to ¾ of its original height, stretched version of the original image to double its width, shortened version of the original image to half its width, shortened version of the original image to ¾ of its original width. (**B**) Same as (**A**) for AM population spiking response. Error bars depict the standard error of the mean (SEM). Asterisks indicate significant differences between familiar and unfamiliar face images for each condition. Shaded area depict confidence intervals (CIs) obtained by bootstrap of the response of the original preferred familiar and unfamiliar faces. Significant differences between them and the distorted versions can be assessed by nonoverlapping CIs.*


A
TP n=16
Response (spikes/s - baseline)
20
10
0
*
*
*
*
original
2H
H/2
3/4H
2W
W/2
3/4W
B
AM n=12
Response (spikes/s - baseline)
10
5
0
original
2H
H/2
3/4H
2W
W/2
3/4W

Internal features and familiar face representations in TP

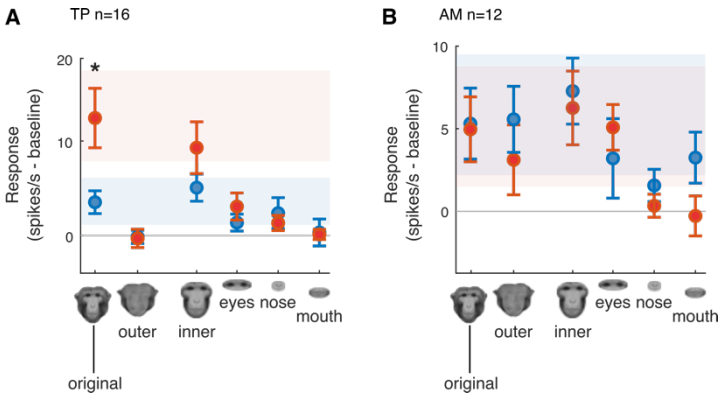
A qualitative difference between familiar and unfamiliar face recognition is that the former depends more strongly on internal (eyes, nose, mouth) rather than external features (ears, hairline) while the latter relies on external rather than internal features (*98-101*). Given that most TP cells encode familiar faces, they could be especially sensitive to presence/absence of internal features.

To test this, five images were created by cropping the original preferred faces: the combined external features (ears and head outline, internal features are absent), the combined internal features (mouth, nose, eyes together – ellipse shape, external features are absent), the eyes, the nose, and the mouth. In TP, only the combined internal features prompted a response that did not differ significantly from the response to the whole preferred familiar face (Figure 15A). In AM, combined internal and external features, as well as the isolated eyes elicited the same response as the original face images (Figure 15B).

TP cells are thus specifically sensitive to the presence/absence of all internal features, and these features presented in isolation induce a significantly smaller response than when presented together. In AM cells are not sensitive to the presence/absence of internal features.

Figure 15. TP cells rely on internal features to recognize familiar faces

*(**A**) Average TP population spiking response to the original preferred familiar and unfamiliar faces of each cell and to cropped parts of them. From left to right, responses to the following images are shown: the original entire face, the outer features, the eyes, the inner features (eyes, nose, mouth), the nose and the mouth. (**B**) Same as (**A**) for AM population spiking response. Error bars depict the standard error of the mean (SEM). Asterisks indicate significant differences between familiar and unfamiliar face images for each condition. Shaded areas depict confidence intervals (CIs) obtained by bootstrap of the response of the original preferred familiar and unfamiliar faces. Significant differences between them and the cropped versions can be assessed by nonoverlapping CIs.*


A
TP n=16
Response
(spikes/s - baseline)
20
10
0
*
original
outer
inner
eyes
nose
mouth
B
AM n=12
Response
(spikes/s - baseline)
10
5
0
original
outer
inner
eyes
nose
mouth

View-(in)variance for monkey faces in TP and AM's neural responses

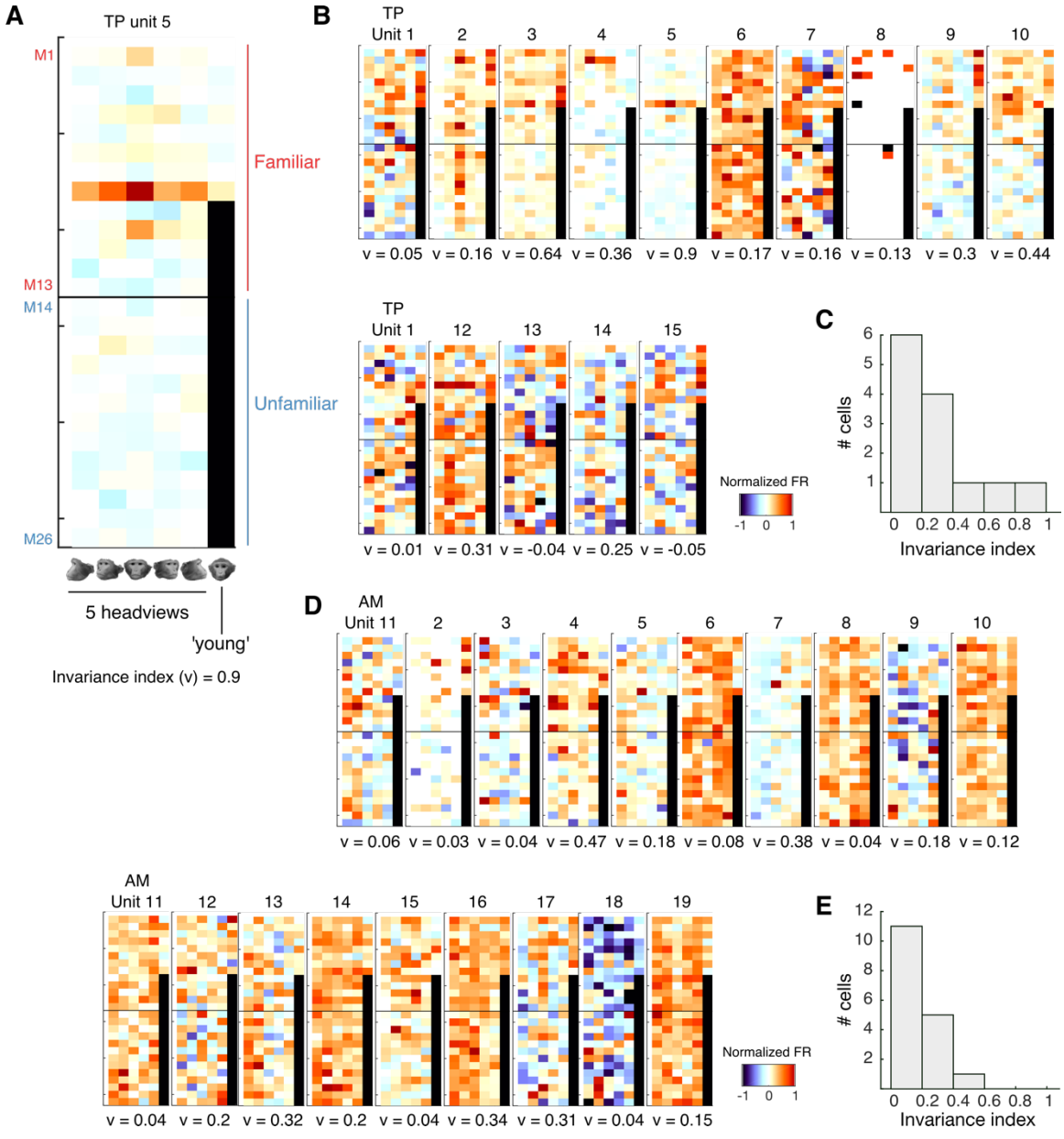
The nature of the stored representation of familiar facial identities is still debated. One theoretical position is that our representation is akin to a prototype which is developed and refined over successive viewings of that individual's face. This prototype could be stored as a 3D model of each identity, thus allowing recognition across different views (*1*). An alternative proposal is that identity representations comprise a series of stored examples of an individual's face (*106*, *107*). Recognition in this case can be achieved when a perceived face is a close match to a stored example, that could comprise a different expression, age and view ('example') of the face. While responses of some cells in AM are invariant to significant changes in viewpoint of human unfamiliar faces (*18*), it is not known how they respond to different views of monkey faces, and whether neural representations are more invariant for familiar than unfamiliar faces. Comparing the response of AM and TP cells to different views of familiar faces will help in understanding how the identity of familiar faces is represented at a neuronal level.

To compare how facial identity is represented across viewpoints in TP and AM, we recorded neural responses of face-selective cells (n=15 in TP, n=19 in AM) to a 130-image set consisting of 13 pictures of familiar and unfamiliar monkey faces in 5 different views (left profile, left half-profile, frontal, right profile, and right half-profile). We also included pictures of 8 of the familiar monkeys that were taken years ago, thus a picture of the 'young' familiar monkeys. An example cell recorded in TP is shown in Figure 16A. This cell responded strongly to familiar monkey identity #8, across all views, and also for the 'young' frontal face picture. We computed view-invariant identity selectivity as a view-invariance index for each cell (see Experimental procedures). TP cells spanned a wide

range of identity selectivity from complete lack of identity tuning (e.g. Figure 16B, unit 6) to sharp identity tuning (e.g. Figure 15B, units 4 and 5). TP cells also spanned a range of view-invariance from view-specific response patterns (e.g. Figure 16B unit 2) to view-invariant responses (e.g. Figure 15B units 3, 4, 5, 10, 12). In AM we did not find any sparse, individual-specific view-invariant response like the example cell of TP shown in Figure 15A. AM cells varied in identity selectivity and view-tuning, and no evidence for a strong view-invariance was found (Figure 16C) for familiar or unfamiliar faces. Thus, we only found a few cells in TP that represented familiar monkey faces in a view-independent manner and we could not reproduce previous results of view-invariance in AM cells.

Figure 16. Head orientation tuning

*(**A**) Response matrix of example cell in TP. Each row shows the responses of the cell to different images of the same identity. Each column shows the responses of the cell to different identities in the same head orientation. From left to right: left profile, left half-profile, frontal view, right half-profile, right profile. The first 13 identities were faces of monkeys personally familiar to the subjects, while the last 13 were personally unfamiliar. The color coding indicates normalized firing rate to the maximum response of the cell. (**B**) Same as (**A**) for all cells recorded in TP. The view-invariance index (v) is indicated for each unit. View-invariance index close to 0 indicates no view invariance to identity. (**C**) Distribution of invariance index for TP cells. (**D**) Same as (A) for all cells recorded in AM. (**E**) Same as (C) for all cells recorded in AM.*


A
TP unit 5
M1
M13
M14
M26
Familiar
Unfamiliar
5 headviews
'young'
Invariance index (v) = 0.9
B
TP
Unit 1
2
3
4
5
6
7
8
9
10
v = 0.05
v = 0.16
v = 0.64
v = 0.36
v = 0.9
v = 0.17
v = 0.16
v = 0.13
v = 0.3
v = 0.44
TP
Unit 1
12
13
14
15
v = 0.01
v = 0.31
v = -0.04
v = 0.25
v = -0.05
Normalized FR
-1
0
1
C
cells
Invariance index
D
AM
Unit 11
2
3
4
5
6
7
8
9
10
v = 0.06
v = 0.03
v = 0.04
v = 0.47
v = 0.18
v = 0.08
v = 0.38
v = 0.04
v = 0.18
v = 0.12
AM
Unit 11
12
13
14
15
16
17
18
19
v = 0.04
v = 0.2
v = 0.32
v = 0.2
v = 0.04
v = 0.34
v = 0.31
v = 0.04
v = 0.15
Normalized FR
E
cells
Invariance index

Beyond faces: individual recognition in the temporal lobe

The neural representation of the whole individual: integrating faces and bodies

Our results so far indicate that TP is involved in the recognition of familiar individuals through their faces. However, recognition can also be achieved by identifying a body shape as belonging to one individual. In some face areas, the response to faces has been found to be enhanced by the presence of an anatomically correctly placed body, and this enhancement grows stronger from posterior to anterior face areas (*76*, *108*, *109*). TP's neural population shows a strong preference for familiar monkey faces, however, some TP cells also responded to unfamiliar monkey bodies (see example cell 2 in Figure 7B, example cell 3 in Figure 10A, cells #16, 24, 25, 30, 49 in Figure 8A, cells #6, 8, 9, 20, 23, 25, 28, 30 in Figure 10C). Given that some cells in TP were sparsely responding to one or a few familiar faces, we wondered whether they would also be selective to familiar bodies. If so, learning whether TP represents the combination of these stimuli, i.e. whole individuals, and how it achieves this would address a larger question of how facial perception transitions into a machinery for familiar individual recognition.

We recorded from TP face-selective cells to determine how they responded to familiar faces and familiar bodies alone, and to whole familiar monkeys. To determine whether TP cells are sensitive to the identity of matching faces and bodies, we created a category of whole familiar monkey stimuli with faces and bodies belonging to different familiar individuals ('incongruent whole-body') (see Figure 17A for stimuli examples).

Figure 17. Putting the face and body together: neural representations of whole individuals

*(**A**) Example stimuli: pictures of four whole familiar individuals ('congruent whole body') were cropped to obtained images of their face, body and also to combine faces and bodies of different identities ('incongruent whole body') (**B**) and (**C**) PSTH of an example cells in TP and AM, respectively. Each row shows the PSTH for one image that was presented for 200 ms, followed by a 500 ms of gray background. The color coding indicates the mean firing rate of 8-12 repetitions of each image in Hz. (**D**) and (**E**) Average population spiking response to all categories in TP and AM, respectively, following the color coding indicated in A. The color shading indicates the standard error of the mean (SEM) for each category. The black line depicts the sum of the measured response to lone bodies and the measured response to lone faces, not the response to an independent stimulus condition. The gray shaded area indicates time intervals in the mean spiking response during which the response to whole bodies (either congruent or incongruent) is supra-additive (estimated using a bootstrap procedure, n=1000 iterations, $p<0.05$).*

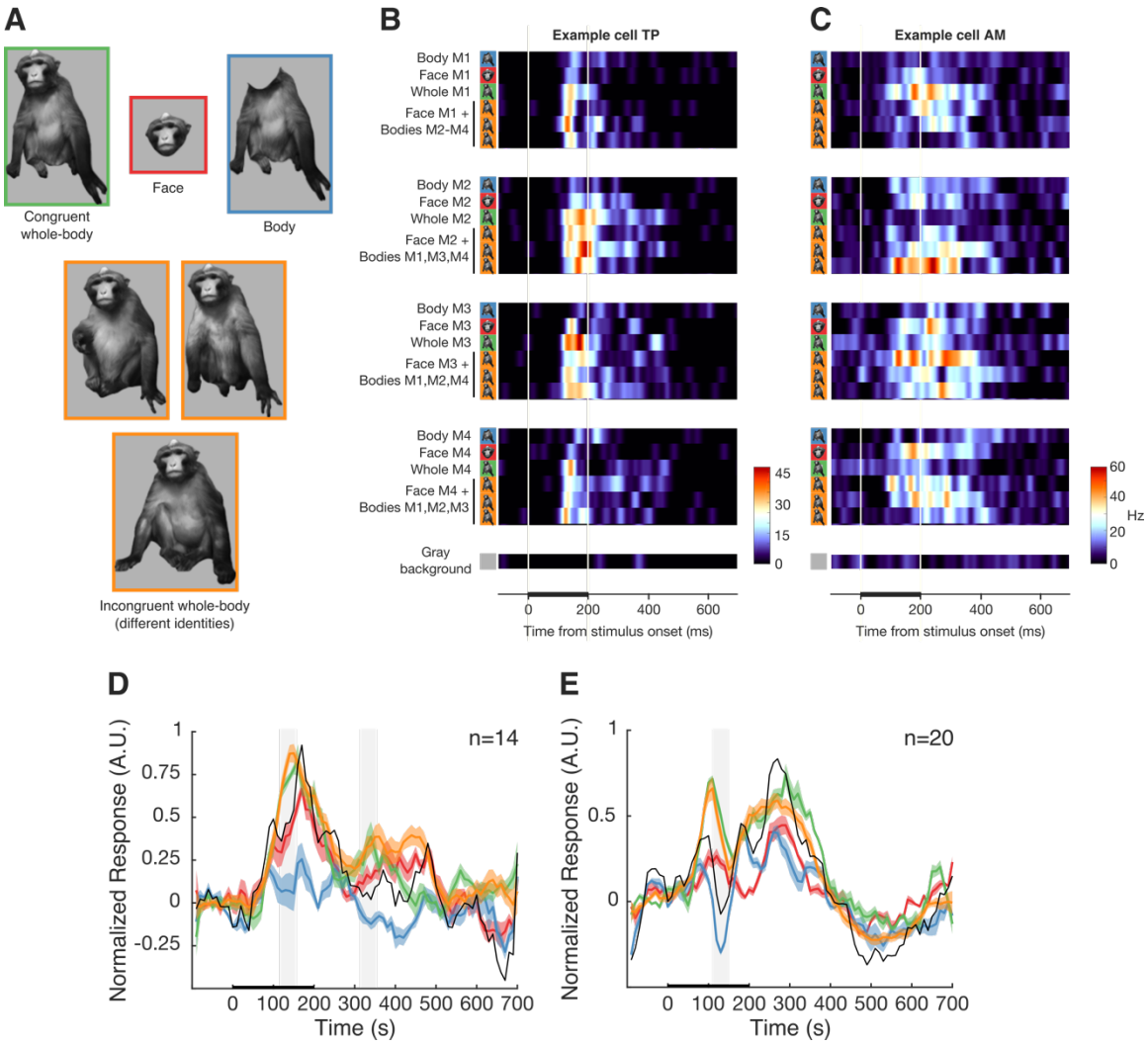

A
Congruent whole-body
Face
Body
Incongruent whole-body (different identities)
B
Example cell TP
Body M1
Face M1
Whole M1
Face M1 + Bodies M2-M4
Body M2
Face M2
Whole M2
Face M2 + Bodies M1,M3,M4
Body M3
Face M3
Whole M3
Face M3 + Bodies M1,M2,M4
Body M4
Face M4
Whole M4
Face M4 + Bodies M1,M2,M3
Gray background
45
30
15
0
0 200 400 600
Time from stimulus onset (ms)
C
Example cell AM
60
40
20
0
Hz
0 200 400 600
Time from stimulus onset (ms)
D
n=14
Normalized Response (A.U.)
1
0.75
0.5
0.25
0
-0.25
0 100 200 300 400 500 600 700
Time (s)
E
n=20
Normalized Response (A.U.)
1
0.5
0
0 100 200 300 400 500 600 700
Time (s)

Figure 17B shows the response of a typical cell in TP to the face-body stimulus set. This cell responded strongly to whole monkeys, and to a lesser degree to faces and bodies. The whole body of monkey 3 elicited the highest response, however, the combination of other faces with bodies also elicited a response with a similar strength. Most cells in TP responded in a similar fashion. The main difference between TP and AM was that AM cells did not respond to bodies as strongly as to faces and responses to whole monkeys were either stronger or similar to the responses to faces (see example in Figure 17C). We did not find any TP cell that preferred one body in particular (data not shown). Both TP and AM cells thus responded maximally to familiar whole-body stimuli, and they did not seem to have a strong identity tuning to familiar bodies or whole-bodies.

The whole-body response was also apparent in the population average response of TP and AM, and was either stronger or similar to the response to isolated faces and bodies (Figures 17D and E, respectively). Averaged population responses to congruent and incongruent whole-body stimuli were similar, indicating that cells in both TP and AM are not sensitive to the identity of matching faces and bodies. In TP the spiking response to whole monkeys was stronger than the sum of the responses to isolated faces and isolated bodies at 120 ms post-stimulus onset (Figure 17D), thus combining in a supra-additive way. AM exhibited a supra-additive combination at a similar latency (110 ms post-stimulus onset). On average, both TP and AM cells responded not only to the sight of faces, but also to the simultaneous observation of faces and bodies, integrating information from both into a unified representation of familiar whole agents.

Identity across sensory modality: integrating faces and voices

Given its anatomical connectivity with both visual and auditory processing streams in the anterior temporal lobe (*110-112*), the temporal pole has been proposed as an ideal candidate for multimodal sensory convergence (*92*, *113*). If this was true, area TP could be comprised of voice cells mixed with face cells, or of cells that respond to both faces and voices, or a combination of these. However, whether cells respond to both visual and auditory stimuli in this region is unknown. Thus, TP provides a unique opportunity to determine how voices impact the representation of faces – and vice versa – within the context of individual recognition.

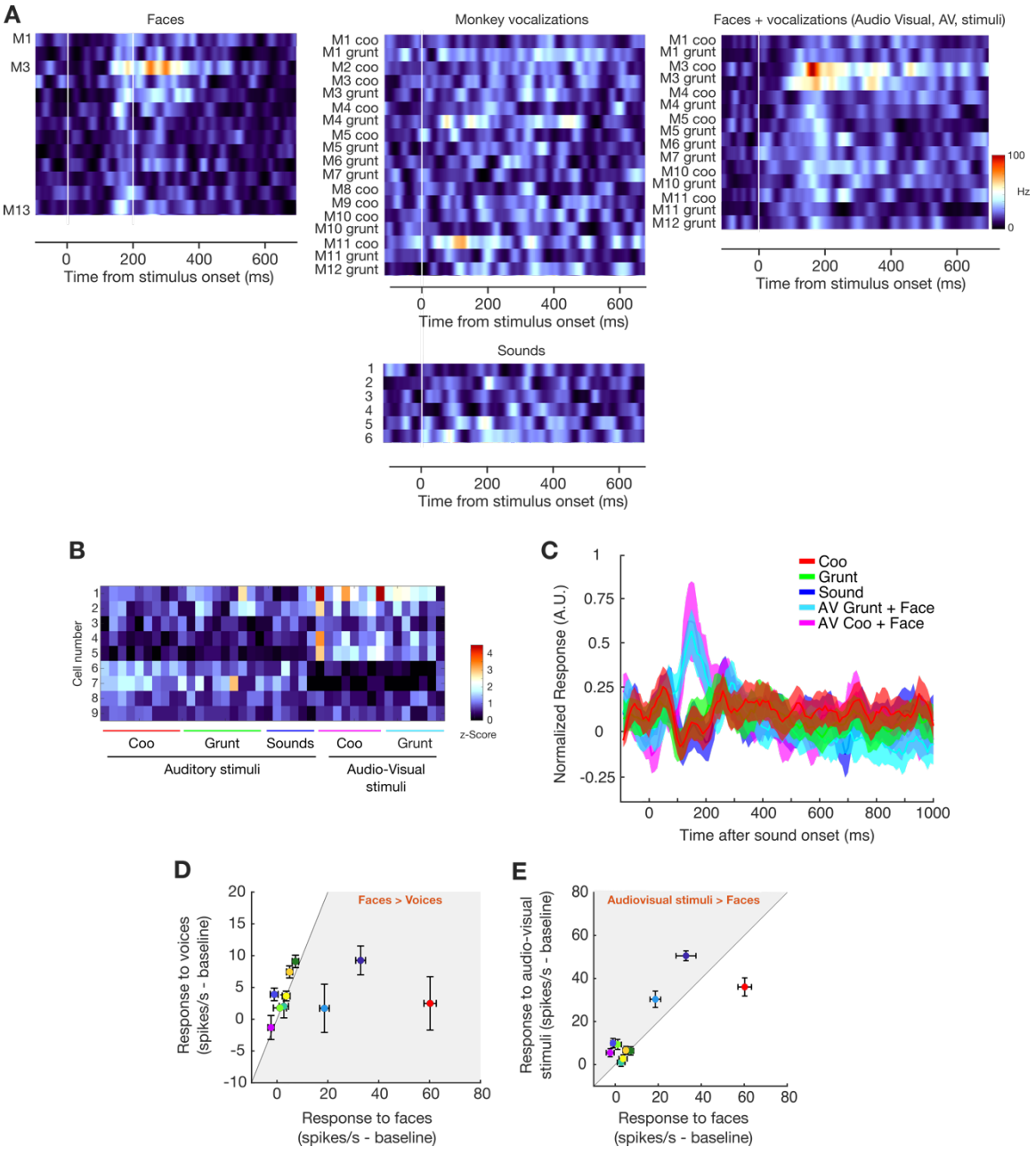
We recorded from nine TP cells to determine if (1) they respond to auditory stimuli, (2) they are selective for voices over other sounds, (3) they respond to both auditory and visual stimuli, and (4) they integrate faces and voices at a single-cell level. An example cell is shown in Figure 18A. This cell responded to some vocalizations (M4 grunt, M10 grunt), to some faces, in particular to the face of M3, and maximally to the audiovisual stimuli of M3 combined face + voice (note the difference in maximum firing rate for these conditions). Different cells varied in their selectivity for auditory and audio-visual stimuli (Figure 18B), but 7 of the 9 cells responded significantly at least to one of these stimuli. Even though for some cells auditory stimuli alone were sufficient to elicit a significant response, on average there was a significant response only to audio-visual stimuli (Figure 18C). When comparing the responses to visual and auditory stimuli, two different types of responses were observed: 3 of the 9 cells were highly visual and 5 responded equally to visual and auditory stimuli (1 cell did not show significant responses to any of the stimuli, Figure 18D). Cells also demonstrated different types of multisensory influences. Some of

them were characterized by audiovisual responses that were significantly higher than the response to faces (Figure 18E, cells plotted above "Audiovisual Stimuli = Faces" line). To sum up, within TP our preliminary results show a considerable sensitivity to several auditory features such as voice identity as well as cross-modal influences on spiking activity.

Codes for familiar individuals appear to be present in the temporal pole: familiar faces, whole-bodies and voices seemed to have colonized the processing resources. However, cognitive maps of visual rather auditory stimuli predominate, probably participating in the read-out of individuals' visual characteristics rather than a vocal. Consistent with the fact that the dorsolateral region of the temporal pole receives projections from third-order auditory association cortex (*110-112*), we encountered clear auditory responses when lowering the electrode in the area dorsal to TP (data not shown). The auditory stimuli were not the best quality, and there was no measure of behavior to parallel and confirm that the subjects were actually recognizing voices (or faces). Auditory stimuli nevertheless influenced visual responses, but it did not seem to do it in an identity-dependent manner.

Figure 18. Individual recognition across modalities: integrating faces and voices

*(**A**) PSTH of an example cell in TP. Each row shows the PSTH for one face (left panel), voice (upper middle panel), sound (lower middle panel) or audiovisual stimulus (AV, right panel) that was presented for the entire length of the auditory stimulus. The color coding indicates the mean firing rate of 8-12 repetitions of each stimulus in Hz. (**B**) Population response matrices in TP (n=9). Responses are sorted from top to bottom by the face selectivity indices (FSI, see Experimental procedures), and from left to right by stimulus category. (**C**) Average TP spiking response to all categories. The color shading indicates the standard error of the mean (SEM) for each category. (**D**) Comparison of visual (faces) and auditory (voices) responses. Each cell is represented by a different color. The line indicates hypothetical equal visual and auditory responses. Error bars depict the SEM. (**E**) Comparison of measured face and audiovisual responses. The line indicates hypothetical equal responses to faces and audiovisual stimuli. Each cell is represented by a different color, which is the same color used in (D). Error bars depict the SEM.*


A
Faces
M1
M3
M13
0
200
400
600
Time from stimulus onset (ms)
Monkey vocalizations
M1 coo
M1 grunt
M2 coo
M3 coo
M3 grunt
M4 coo
M4 grunt
M5 coo
M5 grunt
M6 grunt
M7 grunt
M8 coo
M9 coo
M10 coo
M10 grunt
M11 coo
M11 grunt
M12 grunt
Time from stimulus onset (ms)
Faces + vocalizations (Audio Visual, AV, stimuli)
M1 coo
M1 grunt
M3 coo
M3 grunt
M4 coo
M4 grunt
M5 coo
M5 grunt
M6 grunt
M7 grunt
M10 coo
M10 grunt
M11 coo
M11 grunt
M12 grunt
100
Hz
0
Time from stimulus onset (ms)
Sounds
1
2
3
4
5
6
Time from stimulus onset (ms)
B
Cell number
z-Score
Coo
Grunt
Sounds
Coo
Grunt
Auditory stimuli
Audio-Visual stimuli
C
Normalized Response (A.U.)
Coo
Grunt
Sound
AV Grunt + Face
AV Coo + Face
Time after sound onset (ms)
D
Faces > Voices
Response to voices (spikes/s - baseline)
Response to faces (spikes/s - baseline)
E
Audiovisual stimuli > Faces
Response to audio-visual stimuli (spikes/s - baseline)
Response to faces (spikes/s - baseline)

Chapter IV.

Discussion and conclusions

Recognizing the individuals we know well, such as a family member or an old friend, is extremely important for social animals like us. Faces play a key role in this process. We can easily close our eyes and form a mental image of a friend's face, and even just a glimpse at the face of a famous person is enough to retrieve what we remember about them. So, familiarity with faces is important to the brain, but *where* is it and *how* is it represented?

In Chapter II, we started answering these questions at a systems level. Using whole-brain functional magnetic resonance imaging, we found that in comparison to unfamiliar faces, the entire face-processing network showed increased activity in response to faces of long-time personal acquaintances. In contrast, faces that were only visually familiar caused a reduction of activity in some areas. Thus, the face-processing network distinguished personally familiar faces from visually familiar faces, explaining for example why seeing a face in an advertisement a hundred times is very different from looking at pictures of our friend's wedding.

Personally familiar faces also recruited two previously unknown face areas at anatomically conserved locations within perirhinal cortex and the temporal pole. These two areas, but not the core face-processing network, responded to familiar faces emerging from a blur with a nonlinear surge, akin to the abruptness of familiar face recognition. Thus, we

found that two temporal lobe areas extend the core face-processing network into a familiar face-recognition system.

In Chapter III, we explored how familiar individuals are represented at a single-cell and population level in the temporal pole face area. Using fMRI guided electrophysiology, we tested the hypothesis that the temporal pole (TP) is an area where the recognition of familiar individuals takes place. We found that TP contained a high degree of neurons selective for familiar monkey faces and several hallmarks of familiar face recognition were present in the tuning properties of these cells. Within TP, we also tested responses to monkey bodies and monkey vocalizations. The maximum activity was elicited by faces and bodies presented together, and audiovisual interactions were evident in some TP neurons. However, faces modulated the responses of TP neurons with greater specificity than bodies or voices. Together, these results reveal the neuronal processes sub-serving familiar face-sensitive fMRI-related activity in primates, generate new hypotheses about the neural representations of person concepts, and clarify the position of TP within the recognition circuitry.

Familiarity representations across the primate brain

Capitalizing on the strength of whole-brain imaging to allow for unbiased comparisons across multiple brain regions, our fMRI results extend and build upon previous findings on familiarity effects. Familiarity modulated face processing in four specific ways (Figure 3E-G and Figure 4). First, visual familiarity generally reduced activity in IT cortex. This result is in agreement with several past electrophysiological results: familiarity reduces neuronal activity overall (*114*), sharpens tuning (*115*), and

enhances response to preferred stimuli (*116*), thereby generating sparser representations. Second, visual familiarity with the stimulus and personal familiarity with the individual differed fundamentally and even resulted in opposite modulations. This difference could be from the massive exposure over years to personally familiar faces, the quality of this exposure (diversity of viewpoints, lighting, expression, as well as distance, depth, color, and motion), or the social relevance and semantics associated to personally familiar individuals (*47*). Third, following past electrophysiological findings of personal familiarity (*117*), which suggested localized representations of familiarity (*60*, *61*, *117*), response enhancement by personal familiarity in our study was limited to face selective areas. Thus the faces that shape face-selective cortex throughout ontogeny appear to alter all the different face representations the face-processing system harbors (*9*, *18*, *76*). Finally, familiarity effects did not grow stronger as face representations are transformed from picture to identity-based formats from posterior to anterior IT core face areas. Instead, personal familiarity with faces modulated the whole face-processing system and prompted the activation of two previously unknown face areas in the perirhinal cortex and in the temporal pole.

Effects of familiarity in putative additional face areas in nearby regions (*118*) and cortical columns (*119*) remains to be determined. To our knowledge, this is the first study to find face areas in the perirhinal cortex and temporal pole in non-human primates. This might be partly due to the technical difficulties in obtaining sufficient fMRI signal in these regions, but also because previous studies used only visually familiar or unfamiliar faces as stimuli. The two novel areas, but not the core face-processing network, responded to

personally familiar faces emerging from a blur with a characteristic nonlinear surge, akin to the abruptness of familiar face recognition (*84*).

The two novel face areas were large enough to be detected with fMRI at highly reproducible, and cytoarchitectonically specific regions of the temporal lobe, in perirhinal cortex and the temporal pole. Similar areas may exist in the human brain, however, higher morphological inter-subject variability and technical difficulties imaging deep temporal lobe areas among other reasons make the precise localization of small functionally specific areas harder than in the rhesus monkey.

How and where face familiarity is encoded has been elusive for decades. Our fMRI results established personal familiarity as a crucial parameter to study in the field of face and object recognition, and identified two areas specifically involved in recognizing familiar faces. These two novel areas thus extend the core face-processing network into a familiar face-recognition system.

Familiar face representation in the temporal lobe at the level of single-cells and populations

Basic properties of neuronal processing

Our fMRI findings pointed towards a strong clustering of physiologically identifiable familiar face-selective cells in face areas TP and PR, extending the idea of domain specificity and modular architecture of the face-processing system into an individual recognition system. We confirmed this with fMRI guided electrophysiological

recordings to face area TP: a significant proportion of TP cells encoded faces of monkeys and, another exhibited highly selective responses to different familiar faces in this category (Figures 7, 8 & 10).

It is interesting that cells in TP are selective for both, the species and the familiarity of faces. Even though our subjects interacted with humans daily, and we included pictures of some of these familiar humans as stimuli, we found no cells selective for them. Moreover, TP's neuronal population did not distinguish between human faces and non-face objects (see Figure 8I), lending weight to the possibility that TP is specialized in the processing of familiar monkey faces. In fact, TP cells behave differently from classic face-selective cells: at the expense of losing differentiation between face and non-face stimuli, they responded with very low spike rates to most faces except to familiar monkey faces. As such, they are not very good face detectors but excellent monkey face discriminators. The high selectivity for specific familiar stimuli is a property that could be used to distinguish individual familiar monkey faces. Given that we targeted electrophysiology recordings with a functional localizer that included only monkey faces, whether a familiar human face-selective area exists in a nearby region remains to be determined.

Because visual face area AM encodes the identity of faces (*17*, *18*), we initially expected it would be particularly selective for familiar faces. However, we did not find evidence for this in our fMRI study: AM activity was modulated by familiarity no differently than other areas in the core face processing system. The fact that we did not find a signature of familiarity in AM with fMRI does not imply that AM cells individually, or intermingled with other cells, do not encode familiar faces. After all, a modular structure, useful for some aspects of object perception (*18*, *120*), might not be a requisite for familiar

face recognition. However, we did not find evidence for this: even though AM population activity patterns separated faces and non-face objects, and also monkey and human faces (Figure 9I), we did not find cells that were specifically tuned to familiar monkey faces (Figures 7, 9 & 11). Individual AM cells thus encode facial identity (*17*, *18*) independently of familiarity. Whether this code interfaces with memory, maybe in the form of percept representations through interconnections with TP and PR, is still unknown.

The nature of familiar face representations

The nature of familiar face representations is debated. One theoretical model proposes that familiar faces are stored as 3D models that get refined over successive viewings of a face (*1*). An alternative proposal is that familiar face representations comprise a series of stored examples of an individual's face (*106*, *107*). Having found cells that respond specifically to familiar faces, we were in an ideal position to shed some light on this matter. While some TP cells achieved almost complete view invariance, responding equally to different viewpoints of the same identity (Figure 16), there was not a clear tendency towards a view-invariant or view-dependent coding across the population of cells. Yet, since the number of cells we tested was low, more recordings are needed to determine whether familiar face representation in TP are based on stored prototypes or multiple exemplars.

Unlike previous studies showing view-invariance for unfamiliar faces in AM (*17*, *18*), the cells we sampled in AM were not view-invariant. In addition, we found several cells that responded to both faces and bodies (Figures 7B, 9A & 11A&C), and the population response to bodies was strong, comparable to the response to faces at some time

points (Figure 9D). One explanation for this is that AM, like the middle face areas (*16*, *121*), might not be topographically homogeneous with regards to its coding properties. Thus, we might have sampled from a small region in which facial identity is not encoded (*122*). An alternative explanation is that we tested pictures of monkey faces', instead of human faces, and that low-level features in these pictures were carefully normalized. In previous studies (*17*, *18*) basic properties of the images such as brightness and contrast were not normalized, and thus they could explain the results obtained: brightness, for example, is usually similar for different views of the same identity than between identities, especially in pictures of human faces. In any case, our preliminary results are not enough for establishing the role of AM's view-invariance in the neural machinery of familiar face recognition. Continuing with these experiments, together with obtaining behavioral data would help in understanding both, whether cross-view matching improves with familiarization of identities and whether changes in tuning of AM neurons support this behavioral improvement.

Hallmarks of familiar face recognition in the temporal pole

Three preliminary results suggest that TP could link sensory inputs with the memory of familiar faces. First, TP cells were specifically involved in identifying familiar faces (Figures 7, 8 & 10), and unfamiliar faces that were physically similar failed to elicit comparable responses in those cells (Figure 12C). It has long been known that physically similar faces are represented by similar patterns of discharges across neurons sampled in monkey anterior IT (*60*). Indeed, a recent study showed that AM cells encode the shape-free aspects of faces in the form of invariant representations (*17*). In our preliminary data,

we found a tendency for physical aspects of faces to impact activity of AM cells (Figure 12C). Through interconnections with TP, these percepts could provide the necessary information to recognize familiar faces.

Second, reproducing our fMRI result, the firing rate of TP cells exhibited a non-linear response to the unblurring of familiar faces (Figure 13A-D). Phase un-scrambling had a similar effect (Figure 13F), but spatial frequencies did not seem to impact TP cells in any special way (Figure 13E). These responses support the role of TP as specialized for familiarity, rather than face detection. In contrast, firing rates of AM cells ramped up with the unblurring of both familiar and unfamiliar faces (Figure 13G), while phase-unscrambling induced a non-linear increment in the response for both familiar and unfamiliar faces (Figure 13I). Unlike the unblurring experiment, in which a (blurred) face is visible in all stimuli, faces were not visible in the stimuli with the highest degree of phase scrambling. Thus, the nonlinearity in AM responses in this paradigm could be signaling the detection of a face, providing another piece of evidence supporting AM as a good face detector, not specialized for familiar face recognition. The lack of a signature for familiar face recognition at the macroscopic level, as detected by fMRI, is matched by the single-cell and population level responses in AM.

Third, while in AM both internal and external features elicited the same responses as the original full-face images, TP responses specifically depended on the presence of internal features (Figure 15). This result provides physiological support for the idea that internal features play a dominant role in the neural representation of familiar faces (*98-101*).

Together, these results suggest that TP cells implement a critical stage in the process of familiar face recognition, emerging either as a 'face-recognition unit' or as a tentative 'identity node' (in Bruce & Young terminology (*1*)) where the recognition of familiar monkeys might take place.

Individual recognition: beyond face perception

Most studies on individual recognition focus on faces. However, in real life we are not surrounded by floating faces: faces are typically attached to a body, they sometimes vocalize and they often move. Learning whether TP represents these additional sources of information, and how it does so address a larger question of how face perception transitions into a system for individual recognition.

The representation of whole individuals in the temporal lobe

Only a few behavioral studies in monkeys have addressed the question of how monkeys perceive the bodies of other individuals (*123*, *124*) or their own bodies (*125*). None have tested whether monkeys can recognize familiar conspecifics from their bodies alone, or whether monkeys rely more on faces or bodies to recognize individuals. TP provides a unique opportunity to re-examine this at the neuronal level: TP cells are strongly tuned to familiar monkey faces and some cells also responded to unfamiliar monkey bodies. Thus, TP cells are an ideal neural substrate to study how bodies impact the representation of faces - and vice versa - within the context of individual recognition of familiar individuals.

In comparison to TP face responses, responses to bodies were weaker and only poorly identity specific. Like humans, who rely more on the face than the body to inform their identity judgements (*82*, *126*), monkeys might also rely more on faces than bodies to recognize familiar individuals.

Faces and bodies have been historically studied separately, and thus only a few studies have looked at the representations of whole individuals in face or body-selective cortex (*9*, *108*, *109*, *127*, *128*). Using fMRI, one of these studies showed that in the macaque face-processing network, responses in the posterior face areas are mainly driven by isolated images of faces, and some anterior face areas (specifically AF) prefer images of a whole-agent (a face atop a body). Extending the preference for whole-monkeys to even more anterior face areas, both AM and TP cells in our study responded more to whole monkeys than predicted by adding response magnitudes to isolated faces with those to isolated bodies. The response to incongruent face-body combinations in TP and AM did not differ significantly from the response to faces atop the body of the same identity, suggesting that the identity of the body might help identification but only through the context it provides. These preliminary results provide a framework to address overarching questions of how the brain combines visual information from distinct but related stimuli to form rich and flexible representations of familiar individuals.

Multi-sensory representations of familiar individuals

Rhesus monkeys can discriminate between vocalizations of conspecifics (*129*) and possess a rich representation of other individuals encompassing both vocal and facial information (*58*, *73*, *130*). The temporal pole, with its dense connections to both visual and

auditory processing streams (*110-112*), emerges as an ideal candidate for multimodal sensory convergence (*92*, *113*) that could hold such rich representations of familiar individuals. If this were the case, the temporal pole could be comprised of voice cells mixed with face cells, or of cells that respond to both faces and voices, or a combination of these. Given that face area TP had been functionally defined with visual stimuli in our fMRI experiments, it constituted an ideal site to test how voices impact the representations of faces, and whether voice-selective cells also exist in this region.

We found that neuronal activity to faces and voices differed significantly in TP. Activity for faces was robust, distinct from that for objects and for the subcategory of familiar monkey faces. In contrast, activity for voices was characterized by small responses that were, at best, weakly identity specific (Figure 18). The reason for this might be that we only tested for auditory responses within a small area functionally defined with visual stimuli. Face area TP might be part of a bigger region that encodes both visual and auditory stimuli, and voice-selective cells might exist in a nearby area. An alternative explanation builds on the past observation that voices are weaker cues than faces for identifying individuals, even when learned in conjunction with faces (*131-133*). Since TP neuronal activity might represent memory recall triggered by sensory cues, this could explain why familiar voices elicited weaker responses than familiar faces, and why the identity selectivity from voices and faces did not match. However, differences in the neural coding for voices compared to faces have also been reported in other studies investigating other brain regions in the macaque. For example, voice-selective cells recorded in a voice-selective area localized with fMRI in monkeys (*134*), in neighboring regions to TP (*135*) are less numerous than face-selective cells found in face-selective areas. Thus, the weak

identity-selective responses to voices compared to faces that we observe in TP might be a perpetuation of different inputs received by TP from these regions.

Even though responses of cells in TP were not as strongly triggered by vocal stimuli as by face stimuli, our preliminary results show evidence for auditory modulation of visual neuronal responses in TP. Given that we only tested vocalizations together with faces of the same identity, we cannot conclude if the modulation is identity-specific. While multisensory integration has been shown in a few studies that recorded from extended regions in the temporal lobe (*62*, *136*), this is the first evidence for face-selective cells in a functionally defined region that are modulated by auditory stimuli. Regardless of the nature of multisensory mechanisms, our preliminary results certainly contribute to the set of regions in the visual cortical processing hierarchy that are influenced by auditory inputs and motivate the hypothesis that identity selective cells have a prominent cross-sensory modulation.

Conclusions

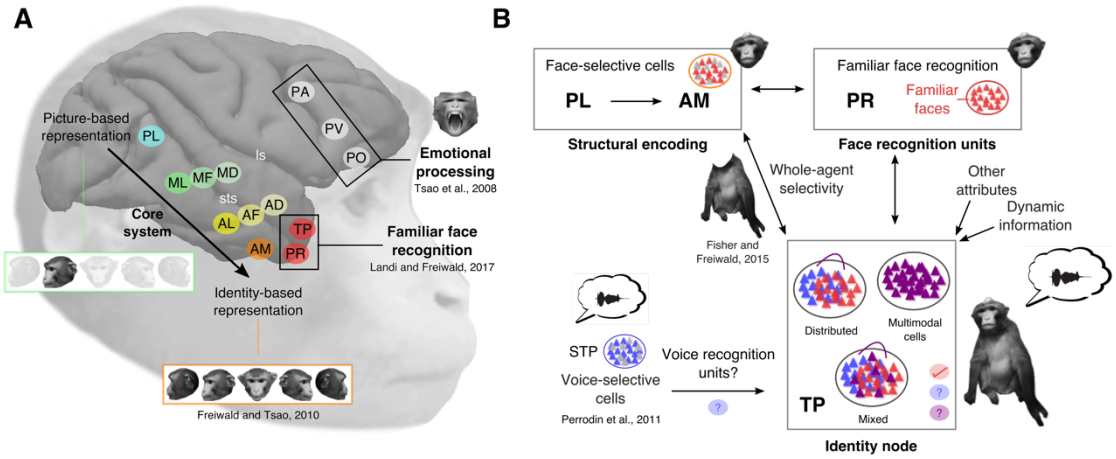
How and where face familiarity is encoded has been elusive for decades. By integrating our results with previous studies, we arrived at a refined model of familiar face recognition (Figure 19). The classic cognitive face-processing model (*1*) postulates a structural encoding system that has been interpreted as the core face-processing network (*48*). In this model, a core face-processing system drives face recognition units using a different coding scheme for familiar faces, which in turn interacts with person identity nodes. We found that personally familiar faces engage these core and extended systems differently. The time course of activation of the novel areas TP and PR conforms to a

pattern predicted for familiar face recognition (*84*) that the core and the extended (prefrontal) face-processing systems lack. This result adds anatomical specificity to earlier models (*1*, *48*). It is tempting to speculate that PR might correspond to the face recognition unit: PR resides in perirhinal cortex, important for declarative memory and perceptual discriminations with high feature ambiguity such as faces (*137*, *138*). On the other hand, the anterior temporal lobe has been suggested as an multimodal 'hub' for individual recognition (*50*, *92*). Most TP cells were selective for familiar monkey faces, and some of them showed signs of multisensory integration, while responses to objects were scarce. These results posit TP as part of a potentially multimodal hub, where associations between different attributes, such as faces, bodies and voices, are processed by common neurons and synapses. Further experiments need to test how nearby regions encode faces and voices of familiar individuals: different modality-specific sub-regions within the temporal pole might process faces and voices independently, and possibly influence each other through direct connections or via an exclusive multimodal region which provides access to individual-specific semantic memory.

Our results suggest two paths from generic face recognition to individual recognition, not one. These two paths from perception to memory are face-specific and not domain-general. Thus, the modular organization of the face processing system is taken at least partly into the memory system, as a specifically face recognition system. At this level, perceptual and mnemonic systems begin to interact to enable the recognition of familiar individuals, and domain-specificity might transition into domain-general representations; as evidenced by selectivity for persons and places in the hippocampus (*28*).

Figure 19. Model of the macaque individual recognition system

*(**A**) Schematic of the macaque face recognition system on the pial surface of the right hemisphere. The core face-processing system performs visual analysis of all faces, regardless of familiarity. The extended prefrontal system has an important role in responding to the emotional content of faces (*11*), but might also be involved in processing low-spatial frequency features contributing to a "first guess" of the face identity. (**B**) Model of individual recognition adapted from Bruce and Young (1986), and complemented with our results. Processing begins with the generation of a view-centered representation of the face, and recognition of a familiar face is based on a structural code involving an abstract facial representation that allows for recognition across changes in pose, expression and illumination. This "structural encoding" occurs in the core face-processing system. Whole-agent selectivity also arises in the core face-processing system (*76*). The identity of familiar individuals is determined further along the hierarchy in face-recognition units and identity nodes. In our results, this property emerges in the extended temporal network in perirhinal cortex (PR) and the temporal pole (TP) face-specific regions. Our results indicate that TP might support multisensory integration of faces and voices, resulting in the activation of semantic information about the individual. Future experiments will determine if seeing or hearing familiar individuals (i) triggers the same temporal pole neurons (multimodal cells), or (ii) if there are different modality-specific sub-regions within the temporal pole that process faces and voices independently and interact through direct connections ('distributed representations') or (iii) via exclusive multimodal cells or a multimodal region which provides access to individual-related semantics ('mixed').*


A
Picture-based representation
PL
ML
MF
MD
sts
ls
AL
AF
AD
AM
TP
PR
PA
PV
PO
Core system
Identity-based representation
Emotional processing
Tsao et al., 2008
Familiar face recognition
Landi and Freiwald, 2017
Freiwald and Tsao, 2010
B
Face-selective cells
PL
AM
Structural encoding
Familiar face recognition
PR
Familiar faces
Face recognition units
Whole-agent selectivity
Other attributes
Dynamic information
Fisher and Freiwald, 2015
Distributed
Multimodal cells
STP
Voice-selective cells
Perrodin et al., 2011
Voice recognition units?
TP
Mixed
Identity node

From face perception to individual recognition: next steps

Using macaque monkeys to study how the brain processes familiar faces has allowed us to rely upon information from both fMRI and electrophysiology experiments, and take advantage of the clear structural organization of their face-processing system. By doing so, we have identified temporal lobe areas where generic face perception transitions to familiar face recognition. However, in order to understand the process of individual recognition from neuron to behavior using this model, four additional steps need to be taken. First, the familiarity effects we observed for TP must be demonstrated at the level of single neurons also in PR. Comparing the activity of neurons in TP, PR, and AM would be the next step to understand how these areas represent familiar individuals. Second, the functional interaction between perceptual and memory processes that underlies individual recognition must be measured. Simultaneous recordings allow for measuring correlations in brain activity such as frequency-specific spike-field coherence (*139-141*), which serve as a proxy for the functional interaction between different brain regions. The physiological interaction between AM, TP and PR renders a plausible neuronal mechanism for individual recognition that flexibly couples perceptual and mnemonic/semantic information. Third, a causal link must be established between the activity of TP and PR neurons and the ability to recognize familiar individuals. While electrophysiology reveals the variety of familiarity signals contained within these areas, neuronal inactivation can be used to understand whether these signals are necessary to recognize familiar individuals. The $GABA_A$ agonist muscimol and optogenetic inhibition techniques have already been used in other face areas to identify their role in face gender discrimination (*142*) and face detection (*143*). By selectively and individually inactivating TP and PR, it will be possible to determine

whether their familiarity signatures underlie behavior during an active recognition task. Finally, the mechanisms underlying recognition through other attributes, such as voices, need to be explored in depth within the temporal pole, and also within the perirhinal cortex. The association between different attributes might represent a basic building block of cognition, shared with other species, and may be an evolutionary precursor of the concept of a person as found in humans.

Final remarks

How the primate brain manages to link perceptual and memory processes is an open question in systems neuroscience. Despite advances in the understanding of face perception and memory alone, the interaction between these two seemingly separate processes has received little attention. The work presented here is a starting point to overcome the gap between visual perception and long-term memory. Using a novel combination of functional imaging with targeted electrophysiological recordings, the experiments in this thesis have laid the groundwork for studies that can address this, advancing the understanding of how facial perception transitions into individual recognition.

Chapter V.

Experimental procedures

All animal procedures complied with the National Institutes of Health Guide for Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees of the Rockefeller University (protocol numbers 12585-H, 15849-H, 18111-H) and Weill Cornell Medical College (protocol number 2010-0029).

Subjects: for the fMRI experiments, we studied four male rhesus monkeys (M1-M4, *Macaca mulatta*, weight 5.8-7.3 kg, 4 years old) that were group- or pair-housed in directly neighboring cages for 2 years before the beginning and for the duration of the experiments. For the electrophysiology experiments, we further studied two of these four monkeys (M1 and M2).

Surgery: a cranial implant composed of MR-compatible acrylic cements (Grip Cement, Caulk; Dentsply International, and Palacos, Heraeus Kulzer GmbH), anchored with MR-compatible ceramic screws (Rogue Research) was implanted in each subject following standard surgical methods, and standard anesthetic, aseptic, and postoperative treatment protocols. A custom-designed MR-compatible headpost (Ultem; General Electric Plastics) were fixed to each implant.

fMRI experiments

Data acquisition

All MRI data were acquired in a 3T MRI scanner while the monkeys were in sphinx position. Functional images were acquired using an AC-88 gradient insert and a custom-designed 4 or 8-channel phase-array receive surface coil with a horizontally oriented single loop transmit coil (H. Kolster, Windmiller Kolster Scientific and L. Wald, MGH Martinos Center, respectively). Each functional time series consisted of 186 gradient-echo echoplanar whole brain images (EPI; repetition time (TR) = 2s; echo time (TE) = 16ms; flip angle [FA] = 80°, 96 x 96 in-plane matrix; voxel size 1x1x1 mm; 54 horizontally oriented slices) acquired in interleaved order with phase partial Fourier 7/8, and two times generalized autocalibrating partially parallel acquisitions (GRAPPA) acceleration.

Immediately before each scanning session, Molday ION was injected into the saphenous vein below the knee to increase the signal-to-noise ratio (SNR). The dose varied from 9 mg of iron per kg of body weight on an initial scan day to 6 mg on subsequent days in adjustment for Molday ION's functional half-life. Molday ION reduces signals in activated voxels, and we thus inverted signals for display of functional data to facilitate visualization.

Anatomical images were collected from each subject in a separate session using a custom-made 1-channel receive coil (L. Wald, MGH Martinos Center) and a T1-weighted magnetization-prepared rapid gradient echo (MPRAGE) sequence (FOV 128 mm, voxel size 0.5 × 0.5 × 0.5 mm, TR = 2.7 s, TE = 2.88 ms, ESP = 6.8 ms, BW = 230 Hz/Px, FA = 9°, 240 sagittal slices) during ketamine and dexmedetomidine (8 mg/kg and 0.01 mg/kg respectively) anesthesia.

Stimuli and procedure

All stimuli were presented at 60 Hz onto a screen placed 36.4 cm from the monkey's eyes using a video projector (resolution 1024 × 768 pixel) with a custom lens. During the experiments, continuous central fixation and pupil diameter were monitored at 120 Hz with an infrared-based eye tracking system. Three block design experiments were conducted with each monkey: a standard face localizer experiment, a Face and Object Familiarity Experiment, and a De-Blurring Experiment. Subjects were trained to maintain fixation on a central fixation spot while in a sphinx position inside a horizontal chair. Importantly, during training subjects were exposed to stimuli from the Face Localizer and from a restricted stimulus set from the Face and Object Familiarity experiments (see below) only. Visual stimulation and reward were controlled using Presentation® software. Fluid reward was delivered after successful fixation for a variable period of time (3.2-5 s), and only runs in which monkeys maintained fixation inside a 2 degrees of visual angle (dva) x 2 dva virtual central window for > 80% of the time were considered for further analysis. Fixation performance was not significantly different across stimulus conditions for any of the four subjects (data not shown, but see Supplementary Information in (*87*)).

Horizontal and vertical eye position, as well as pupil diameter were resampled at 1kHz and stored in Presentation log files. Blinks were removed by linear interpolation of the values measured 100 ms before until 100 ms after the blink. The same procedure was applied to remove any pupil recordings that deviated more than 3 standard deviations from the mean pupil trace (e.g. due to saccades). The interpolated pupil time series were smoothed with a Butterworth third order filter (cutoff 4 Hz). Baseline pupil diameter on each trial was calculated as the average diameter during the 8 seconds of gray background

preceding the block of stimuli. Pupil response was calculated as the change from baseline of the mean pupil diameter, for each 24-s stimulus block: $\boldsymbol{d} = \frac{\boldsymbol{m}-\boldsymbol{b}}{\boldsymbol{b}}$, where $\boldsymbol{b}$ is the corresponding baseline for each block, and $\boldsymbol{m}$ is the mean diameter per category per block. $\boldsymbol{d}$ was averaged for each stimulus category across blocks for each run. A one-way ANOVA with stimulus category as a factor was conducted on the averaged pupil response per run to determine whether pupil response varied between categories. Pupil response was normalized to the maximum response for each subject for plotting purpose (Figure 2D).

Face Localizer Experiment

To localize face-selective areas, we used a standard face localizer (*9*). In short, this experiment consisted of a block design containing blocks (24 s each) of a fixation-only condition and six stimulus categories of grayscale pictures: rhesus macaque faces, cynomolgus macaque faces, body parts and headless bodies, manmade objects, fruits, and phase-scrambled versions of the rhesus macaque faces. Each block contained 15 different images from within a single category, and each individual stimulus was presented 4 times per block for 400 ms in a pseudo-random order. Low-level characteristics across the various categories were equalized: images of faces, fruits, objects and bodies were first cut out from their original background, matched within and across all categories for total screen area, then normalized for luminance and frequency amplitude using the SHINE toolbox (*144*), and finally placed on a pink noise background. The pink noise background subtended 30 dva^2, and the average foreground image was 14.2 dva x 11 dva. Images were presented at five different locations on the screen: at the center, or 2.2 dva from the center to the top, bottom, left or right. A Latin square design was used to balance the order of the

six blocks within a run and the identity of the immediately preceding block. Each run consisted of 36 s of gray background, followed by the 6 stimulus blocks, 24 s of gray background, the same 6 stimulus blocks presented in the reverse order, and concluded with 24 s of gray background. 27, 30, 33, and 32 runs of this experiment were analyzed for monkeys, M1 to M4, respectively.

Face and Object Familiarity Experiment

Three sets of stimuli were created for each subject to examine familiar and unfamiliar face and object recognition: (1) *Personally familiar stimuli*, which contained pictures of familiar individuals and objects with whom/which the subjects interacted daily; (2) *Unfamiliar stimuli*, containing pictures of unknown individuals and objects which the subjects had never seen before; and (3) *Visually familiar stimuli*, containing pictures of individuals and objects unknown to the subject in real life, but which, as pictures, the subjects had seen during training hundreds of times. These sets of stimuli are described in detail below.

Personally familiar stimuli, pictures of personally familiar faces (PFF) and personally familiar objects (PFO): The same four subjects that were studied were shown in the personally familiar face stimulus set. For picture acquisition, a blue screen was placed behind the stimulus subjects under a standardized set of lighting conditions. The monkey bodies were masked out with a blue neck plate, and the faces were isolated from the background. 10 pictures of the face of each monkey with a neutral expression were selected, and for each experimental subject, a stimulus set of the 30 pictures of the faces of the other individuals was compiled.

Similarly, cage toys the monkeys interacted with on a daily basis were shown in the PFO picture set. 10 images containing different views of each toy were created (30 images in total).

Unfamiliar stimuli, non-familiar faces (NFF) and non-familiar objects (NFO): NFF were obtained from the PrimFace database (http://visiome.neuroinf.jp/primface). 10 pictures of unobstructed faces of three male rhesus macaque monkeys of ages 3-5 years (same age and sex as the personally familiar monkeys) with a neutral expression (same as PFF) and view matched to the PFF set were selected (30 images in total). After normalizing all stimuli across low-level characteristics in the same fashion as for the Face Localizer Experiment, there were no low-level or mid-level (e.g. age, sex, pose) noticeable differences between PFF and NFF.

NFO were downloaded from various sources of well-established stimuli used for novel object recognition studies in humans (*145-147*). Three objects were selected, and 10 pictures of each object, including rigid rotations in depth around the vertical axis by 0, 45, and 90 degrees, were included.

Visually familiar stimuli, visually familiar faces (VFF) and visually familiar objects (VFO): Whenever subjects were trained, they were shown a video from a documentary series produced by National Geographic during breaks (http://natgeotv.com/uk/monkey-thieves). 5 pictures of three individuals and three objects were extracted from frames in the movie, cropped and edited to isolate them from the background (15 face images and 15 object images in total). Only pictures of faces with a neutral expression were chosen. This stimulus set, together with the face-localizer stimuli, were used to train subjects on fixation in the weeks preceding the experiment, and thus, images of faces and objects belonging to

this set were visually familiar to the subject. Visually familiar images were seen 392 times by M1, 264 by M2, 520 times by M3, and 200 times by M4, over the course of the 9, 7, 13 and 6 days preceding the experiment. The movie from which we took these images was seen in its totality at least 20 times by M1, 9 times by M2, 11 times by M3, and 9 times by M4. The amount of exposure we used is comparable to other studies that investigate the role of visual familiarity in inferotemporal cortex activity (e.g. (*114*, *148*)).

All pictures were converted to grayscale and low-level characteristics across the various categories were equalized for the Face and Object Familiarity Experiment in the same way as for the Face Localizer Experiment. The pink noise background was presented at 30 dva^2, and the average foreground image was 8 dva x 6 dva. Images were presented at five different locations on the screen: at the center, or 1.1 dva from the center to the top, bottom, left or right. A Latin square design was used to balance the order of the six stimulus categories, resulting in a block design of equal characteristics as the face localizer experimental design.

For this experiment, subjects were scanned 8 days in total. All sessions included pictures of the same face and object identities. PFF, NFF, PFO, and NFO were initially visually novel and became visually familiar over the course of first four days of scanning. On day five these stimuli were changed and were thus visually novel again. VFF and VFO were not changed over the course of the experiments (see Figure 2C). At least 12 runs per day were analyzed for each monkey, and 100, 110, 99 and 99 runs of this experiment were analyzed, respectively for M1 to M4.

De-Blurring Experiment

Images of PFF and NFF used in previous experiments were included in this experiment. The level of visual familiarity for these stimuli was thus equal. A new set of object images that look like faces after low-pass filtering was designed and included. These objects were fruits, vegetables, mushrooms, and butterflies that resemble facial shape and symmetry. The subjects were trained for the De-Blurring Experiment for eight days prior to starting the experiment using these object stimuli. All stimuli were thus visually familiar to the same degree by the time the experiment took place. Images were first matched within and across category for total screen area and normalized for luminance and frequency amplitude using the SHINE toolbox in the same way as with the face localizer stimuli. Images were placed on a gray background with the average foreground image occupying 8 dva x 6 dva and the whole image occupying 30 dva^2. For each stimulus, 80 images with different degrees of blurring were generated and sorted by increasing amounts of high spatial frequency information (starting point: 7 cycles per image (cpi); increments of 1/16 octave per image). Images were concatenated into a 32-second movie (16 TR) at 2.5 frames per second. The HSF content of each individual stimulus therefore increased progressively over the course of each movie from 7 cpi to 36.64 cpi.

A Latin square design was used to create the stimulus blocks order. Within each run the same views of the face were presented for the personally familiar faces and the unfamiliar faces. Each run consisted of 36 s of initial gray background and six stimulus blocks of 32 s each separated by 24 s of gray background intervening gaps. 45, 45, 53, and 51 runs were analyzed for monkeys, M1-M4, respectively.

Data analysis

Data analysis was performed using FREESURFER and FS-FAST software (v5.3, http://surfer.nmr.mgh.harvard.edu/), AFNI scripts (http://afni.nimh.nih.gov), the "JIP" toolkit provided by J. Mandeville (http://www.nitrc.org/projects/jip/), and custom scripts written in MATLAB. Raw functional data from each scan session were corrected for head and body motion with a slice-wise correction allowing for shifts, rotations, scaling, and skewing within each slice and included terms for cubic warping in the phase-encoding direction. Slice-timing correction was applied using FreeSurfer's tools, and JIP's mutual information-based non-linear alignment was used to compensate for static distortions of the functional volumes and to align these volumes to the high-resolution anatomical images that were used for cortical surface modeling. Functional data were analyzed using a general linear model (GLM), with motion parameters included as nuisance regressors. No spatial or temporal smoothing was applied. For the individual GLM fitting, the first five volumes of each functional run were excluded to prevent T_1 saturation artifacts, and whitening (first-order autoregressive model) and detrending (first and second order polynomials) were applied.

To create cortical surface reconstructions, individual anatomical images were segmented, smoothed, and inflated separately for each hemisphere (*149*, *150*).

Whole Brain Analysis

In order to conduct whole brain statistics, an anatomical volume template brain was created in AFNI and aligned to the MNI-Paxinos macaque template brain (*151*, *152*). Because the sample size is only 4 subjects, slightly larger than the typical monkey fMRI

studies (*153*, *154*) but still too small for effective power in a random effects analysis (*155*), a fixed-effects group analysis was used as in previous studies (*9*, *153*, *156*). Fixed-effects group analysis were therefore performed and supplemented with single-subject analysis, and the level of significance was set at $q < 0.05$ (Benjamini-Hochberg false discovery rate -FDR), correcting for multiple comparisons. For the Face and Object Familiarity Experiment, a repeated measures ANOVA with conditions (personally familiar faces, unfamiliar faces, visually familiar faces, personally familiar objects, unfamiliar objects, visually familiar objects) and time (scan day) as factors was carried out.

Region of Interest (ROI) Analysis

Face-selective ROIs: as in previous studies, face areas were labeled based on anatomical location and relative position (*8*, *9*, *18*, *157*) with the Face Localizer Experiment, using the contrast Faces versus Objects (Figure 1). Subsequent Face and Object Familiarity Experiment (Figure 3A) yielded the exact same locations for all face areas. All ROIs were defined with the contrast PFF > PFO by selecting a local maximum signal change in a region and selecting a 3 x 3 x 3 voxel region centered on this point, excluding all voxels that did not meet the criteria $q < 0.05$ (Benjamini-Hochberg FDR). Except for prefrontal face area PA, which was present only in two of the subjects (same frequency as previously reported (*11*)), all face areas were present in the four subjects. All ROIs were defined in each hemisphere and then pooled together into bilateral ROIs. An independent analysis using ROIs defined with the Face Localizer Experiment (contrast Faces versus Objects) yielded the same effects for both the Familiarity and the De-Blurring Experiment. To avoid double dipping, we defined the ROIs used for the Face and Object Familiarity Experiment

with the results of the first four days, and we performed statistical tests with the data of the last four days of the experiment (*158*).

Object-selective ROI: an object ROI ('Obj') was defined with the contrast PFO > PFF in the anterior lip of the STS, approximately 4 mm posterior to the face area AL in keeping with past conventions in the literature (*80*) (Figure 3C).

Holm-Bonferroni corrections for multiple ROIs were used to adjust significance thresholds to control for family-wise error rate (*159*). The same ROIs were used for the De-Blurring Experiment.

For the purpose of making post hoc comparisons, we defined three groups of face-selective grouped ROIs: core network (PL, ML, MF, MD, AL, AF, AD, and AM), prefrontal network (PO, PV, and PA), new anterior temporal ROIs (TP and PR), and one object-selective group (Obj).

For the Face and Object Familiarity Experiment, a three-way ANOVA was conducted to determine the effects of familiarity type (personal versus visual), object category (faces versus object) and grouped ROI on % contrast effect size. A one-way ANOVA was conducted to compare the effect of grouped ROI on each % contrast effect size, and a separate two-way ANOVA was ran to test for interaction effects between stimulus category and type of familiarity, within each grouped ROI.

Time Course Analysis

For the De-Blurring Experiment, ROI time courses were created by averaging the signal from all voxels for each time point within each ROI for individual runs, smoothed with a 3-point moving average, and detrended for first and second order polynomials to

remove signal drift across each run. The percent signal change (PSC) of the fMR signal was calculated for each stimulus block using the last 3 TRs of the previous gray background condition epoch as baseline condition, and averaged across runs for each subject. Activation profiles in each ROI were quantified by fitting the Naka-Rushton sigmoidal function (*86*) to the time courses from stimulus onset until peak activation:

$$\boldsymbol{R(t) = R_{max} * \frac{t^n}{t^n + c^n} + d}$$

where $\boldsymbol{R}$ is the PSC from baseline, $\boldsymbol{R_{max}}$ is the asymptotic PSC, t is time measured as number of TRs, $\boldsymbol{c}$ is the TR at which the response is half of $\boldsymbol{R_{max}}$, $\boldsymbol{d}$ is baseline PSC, and $\boldsymbol{n}$ is the steepness of the response. This function has been used extensively to characterize the response profile of face areas as a function of specific stimulus dimensions, such as facial geometry (*160*). A bootstrapping procedure (*161*) was used to test if there was a significant difference between the parameter values fitted to the three categories' data sets. The 95% confidence intervals (CIs) computed from these bootstrapped samples were used to estimate differences in the curve steepness between conditions for each individual ROI. Significant differences in the parameters were inferred by non-overlapping bootstrapped CIs.

A two-way ANOVA was conducted to compare the effect of stimulus category (familiar faces, unfamiliar faces and objects) and grouped ROI (core, prefrontal, new anterior temporal and object) on response steepness.

fMRI-guided electrophysiology

We used the Planner software (Ohayon and Tsao, 2012) to calculate the angle and position of desired electrode trajectories based on functional and contrast-enhanced anatomical MRIs. This allowed us to highlight blood vessels to be avoided. We implanted an MRI-compatible recording chamber (19-mm diameter, 6-ICO-J2A 35-degree angle in M1, 25 degree in M2, Crist Instruments) under aseptic surgical conditions. Planning and intra-operative guidance of instruments was performed with the neuronavigation system cortEXplore. (developed by Stefan Schaffelhofer). We used plastic grids 3D-printed from plans created by the Planner software (Ohayon and Tsao, 2012) placed inside plastic recording chambers. TP and AM were targeted in the right hemisphere of monkey M1 and M2 (see Figure 7A).

Extracellular recordings

For recordings, tungsten electrodes (1-3 MOhm/1KHz, FHC) were back loaded into metal guide tubes, which also served as the reference. Guide tube length was set to reach from the top of the grid holes, approximately 5-7 mm below the dura. The electrode was slowly advanced using an oil hydraulic micromanipulator (Narishige Scientific Instrument).

The electrophysiological signal was amplified and recorded at 30k Hz. Waveforms that crossed a set threshold were sorted online into separate units using multiple discrimination windows. Spiking activity, local field potentials, eye position traces and digital triggers of task events were recorded using a Cerebus Data-Acquisition System (Blackrock Microsystems). Spikes were detected online based on their waveform and later

sorted offline using Blackrock offline spike sorter (Blackrock Microsystems). We recorded from each single unit encountered along the electrode trajectory within these areas. Each well-defined cluster was treated as a 'cell' for the purposes of the analyses.

On each recording day, the electrode was initially advanced based on MRI-planning until the border of face patch TP or AM was reached. The presence of neural activity was monitored by the sound of the amplified signal connected to an audio monitor and by observing the presence of waveforms on a computer screen. We assessed the consistency of transitions between gray matter, white matter and sulci assessed from these measures with the expected trajectories in anatomical MRIs. Expected visual, auditory, motor or somatosensory responsiveness along the trajectory was verified and compared with that expected from brain atlas (*152*) as well as from preceding recording sessions.

Behavioral monitoring and stimulus presentation

Behavior was controlled and stimuli were presented using custom software 'Visiko' written in C++. This software runs on a Windows PC that received and sent signals via an analog and digital input/output card PCI-DAS1002 (Measurement Computing Corporation). This software controlled the reward and sent triggers to the Cerebus data-acquisition system. Eye position was measured and recorded at 100Hz using an infrared eye-tracking system (ETL- 200, ISCAN Inc.). Stimuli appeared on a CRT (cathode ray tube) computer monitor (36.6 x 27.4 cm; 1920 x1440 pixels; 100 Hz refresh rate) at a distance of 57 cm from the eyes.

During all experiments, monkeys were required to fixate within a 3x3x3 DVA fixation window and were given drops of juice and/or water as reward for fixation performance every 3 to 4.5 s.

Experimental design

In all visual experiments, pictures were presented in random order for 200 ms, separated by 500 ms blank interval which consisted of a gray background. Low-level characteristics across the various pictures were equalized: images were first cut out from their original background, matched within and across all categories for total screen area, then normalized for luminance amplitude using the SHINE toolbox (*144*), and finally placed on a gray background.

For all experiments, only the stimulus presentations for which fixation was maintained 100 ms before the presentation, throughout the presentation, and after 100 ms after the presentation were considered for further analysis (a total of 400 ms). The order of experiments was not randomized, and the experimenter was not blinded to the experimental conditions. After completing experiments, we advanced the electrode until the next cell was reached.

Face and Familiarity selectivity

After a unit had been isolated, we first determined single-unit, multi-unit, and evoked potential selectivity to 5-12 repetitions of an image set consisting of 95 grayscale images. These set consisted of human faces (13 personally familiar, 13 unfamiliar), monkey faces (12 personally familiar, 13 unfamiliar, 1 image of himself), bodies (13

unfamiliar bodies), objects (13 personally familiar, 13 unfamiliar), and a gray background (4 images) (Figure 7B). The pictures of faces consisted only of frontal views.

Extension of Face and Familiarity Experiment

This experiment was run in each site in which face-selectivity was established during the initial screening at either single-unit, multi-unit or evoked potential level. In this experiment, neural responses were monitored while the monkey was looking at an expanded 205-image set which included more items of some of the screening categories: human faces (30 personally familiar, 30 unfamiliar), monkey faces (12 personally familiar, 72 unfamiliar, 1 image of himself), bodies (15 unfamiliar bodies), objects (15 personally familiar, 25 unfamiliar), and a gray background (5 images). The pictures of faces consisted only of frontal views. Depending on single-unit stability, we run 8-12 repetitions of this stimuli set. If the single-unit was stable by the end of this experiment, we run further experiments. We used this experiment to determine the preferred familiar and unfamiliar face for each unit.

Physical similarity measurements

A set of 70 landmarks were labeled on each of the monkey frontal face images. 42 of the landmarks were placed in "inner" face features and 28 landmarks were placed in the borders of the face, as "outer" features (Figure 12A). The positions of landmarks were normalized for mean and variance for each of the 85 faces, and an average shape template was calculated. Then each face was smoothly warped so that the landmarks matched this shape template, using a technique based on bilinear interpolation. This warped image was

then normalized for mean and variance, resulting in a 'shape-free' image. Principal component analysis (PCA) was carried out on positions of landmarks ('shape') and shape-free intensity ('appearance') independently to determine if face space could be explained by a few dimensions.

Further selectivity and invariance to the preferred faces

To test a cell further, it had to comply the following criteria: (1) face selectivity, defined by the Face Selectivity Index (FSI, see below), (2) familiar or unfamiliar monkey faces as preferred category (see below), and (3) Percentage of stimuli > half-maximum response less than 5% (see below). We tested cells with these characteristics with different hallmarks of familiar face recognition. For each cell, we created the following stimuli from the familiar and unfamiliar face that elicited the highest response:

Un-blurring: for each preferred face, 10 images with different degrees of blurring were generated by increasing amounts of high spatial frequency information (high pass filter ranged from 5 cycles/face to full spectrum, Figure 13D & G).

Frequency decompositions: four spatial frequency band images were created from the original preferred faces (Figure 13E & H). Partial frequency content in the original stimuli was filtered using a high-pass cut-off that was > 20 cycles/face for the high spatial frequency stimuli (HSF), and a low-pass cut-off that was < 5 cycles/face for the low spatial frequency (LSF) stimuli. Two intermediate broad-band frequency (BSF) images were created, one containing 10-20 cycles/face (high BSF) and one containing lower frequencies 5-10 cycles/face (low BSF).

Un-phase Scrambling: for each preferred face, we incrementally scrambled the Fourier phase at 10 distinct levels, while leaving the amplitude spectrum unaltered (Figure 13F & I). Faces were scrambled by computing a two-dimensional fast Fourier transform (FFT), which yields a complex representation (amplitude and phase). The phase values were then randomized by assigning a random value to each element taken from a uniform distribution across the range. The proportion of the phase values that were randomized increased from 0% (the original image) to 100% (fully phase scrambled) in 10 steps. An inverse FFT was applied to the resulting amplitude/phase maps to produce the scrambled version.

Distortions of face geometry: we created six images from the preferred faces changing the vertical and horizontal dimensions independently (Figure 14). Vertical (horizontal) stretch: two images were produced by vertically (horizontally) stretching the normal faces to x2 and x1.5 their original height (width) while preserving their original width (height). Vertical (horizontal) stretch: two images were produced by vertically (horizontally) stretching the normal faces to x2 and x1.5 their original height (width) while preserving their original width (height). Vertical (horizontal) shrink: one image was produced by vertically (horizontally) shrinking the normal faces to x0.5 their original height (width) while preserving their original width (height).

Facial features: Five images were created by cropping the original preferred faces: (1) the external features (ears and head outline – ellipse shape containing inner features empty), (2) the internal features (mouth, nose, eyes together – ellipse shape, no external features), (3) the eyes, (4) the nose, (5) the mouth (Figure 15).

Invariance experiment

For some cells, we tested invariance to head orientation and to changes induced by age in monkey faces (Figure 16). The image set contained pictures of 26 individuals in each of 5 head orientations (left full profile, left half profile, frontal, right half profile, right full profile). Of these 26 individuals, 12 were familiar and 13 unfamiliar to the monkeys, and one individual was the subject himself. The frontal view of these individuals was the same picture as the one used in the Face and Familiarity selectivity experiment. We also included pictures of 8 of the familiar monkeys taken when 3-5 years before the experiments took place ('young' pictures of familiar individuals). This experiment was run on cells that were determined to be face selective by their face selectivity index (FSI, see below).

Face-body associations

The stimuli set for this experiment consisted of pictures of 4 familiar individuals. These pictures included their face, their head-less body or the face and the body ('congruent whole body'). Pictures of faces and bodies of different individuals were combined to create an incongruent category of 'incongruent whole-body' (Figure 17A). This experiment was run on units that were determined to be face or body selective by the initial screening.

Face-voice associations

The stimuli set for this experiment consisted of vocalizations of 11 familiar monkeys, 6 unfamiliar monkeys and sounds of 6 objects. The mean sound pressure level for the duration of each audio sample was calculated. Then the audio samples were normalized for this mean acoustic intensity (MATLAB, Math- Works, Natick, MA).

Auditory stimuli consisted of recordings of 2 s duration, each containing either a sample of "coo" vocalization (739 ± 230 ms), a "grunt" vocalization (599 ± 311 ms) or a sound (1560 ± 522 ms). The vocalizations and sounds were played every 2 seconds, alone ('Auditory condition') and while vocalizations were also played together with the presentation of a static picture of the face corresponding to the familiar individual ('Audiovisual condition'). This experiment was run on cells that were not necessarily face-selective but that were located in TP according to anatomical and functional landmarks of surrounding tissue.

Analysis of neural responses

The data were analyzed with custom written scripts and the Statistics Toolbox in Matlab (MathWorks). Spike times were saved at a resolution of 1ms. Peri-stimulus time histograms (PSTHs) were obtained using bins of 10 ms Gaussian smoothing (full-width at half-maximum, FWHM). Neurons with very low firing rates (FRs) (<0.5 Hz) were not included in the analyses. The baseline activity was defined as the average activity during 100 ms preceding the stimulus onset. The cell's response to a given stimulus was defined as the trial-averaged firing rate (FR) in a 200-ms window centered on the time of the peak stimulus activity. In parallel, spike trains were also smoothed by convolution with a Gaussian kernel (σ = 10 ms) to obtain the spike density function (SDF). Cells were considered visually responsive if they gave a response that was significantly different from baseline activity ($p<0.05$, t-test), and if its SDF was greater than the mean plus 4 standard deviations (SDs) of the baseline activity (*62*).

Throughout the analyses, to compute the mean normalized response to a stimulus, for each cell, the mean baseline response was subtracted, and the responses to all stimuli were normalized by the maximum response. To compute Z-scores, the average response to each stimulus was normalized to standard-deviation units (SD) with respect to baseline.

"Face-selective" cells were classified according to the Face Selectivity Index (FSI) criterion used in previous studies (*23*). The FSI was defined by

$$FSI = \frac{mean_{faces} - mean_{others}}{mean_{faces} + mean_{others}}$$

where $mean_{faces}$ was the mean response above baseline to all face stimuli and $mean_{others}$ the mean response above baseline to non-face objects. An FSI of 0 indicates equal responses to face and non-face objects. An FSI of 0.33 indicated twice as strong response to faces as to non-face objects. We established this threshold for determining whether visually responsive cells were face-selective or not. For cases where ($mean_{faces} > 0$) and ($mean_{others} < 0$), FSI was set to 1; for cases where ($mean_{faces} < 0$) and ($mean_{others} > 0$), FSI was set to -1.

Face familiarity indexes were calculated in a similar way to the FSI. We calculated separately an index for human and for monkey faces, defined by:

$$FFI = \frac{mean_{familiar\ faces} - mean_{unfamiliar\ faces}}{mean_{familiar\ faces} + mean_{unfamiliar\ faces}}$$

Neurons were considered modulated by familiarity in the cases where FFI>0.25.

The preferred category for each cell was defined based on the stimulus that elicited the largest (maximal) response amplitude.

Similarity matrices of correlation coefficients were computed between the population response vectors (across all cells, averaged over stimulus repeats) to each of the stimuli in the 95-image set.

Classical multi-dimensional scaling was performed on the population responses to the 95-image set of all cells in each area, using a Euclidean distance metric.

For each recording site, a stimulus-dependent local-field potential map was computed by averaging the evoked LFP response to repeated presentations of each stimulus; the map was normalized by subtracting the mean.

For face-selective each cell, sparseness was defined as $S = \frac{(\sum_{i=1}^{n} \frac{r_i}{n})^2}{\sum_{i=1}^{n} \frac{r_i^2}{n}}$

with r_i being the baseline corrected response of the neuron to the ith face and n being the total number of faces.

The view-invariance index was defined as the mean correlation between the 26-element response vector at the preferred head orientation and the 5 response vectors at the remaining head orientations (*18*).

References

1. V. Bruce, A. Young, Understanding face recognition. *British Journal of Psychology*. **77**, 305–327 (1986).

2. H. D. Ellis, D. M. Jones, N. Mosdell, Intra- and inter-modal repetition priming of familiar faces and voices. *Br J Psychol*. **88 (Pt 1)**, 143–156 (1997).

3. S. Campanella, P. Belin, Integrating face and voice in person perception. *Trends in Cognitive Sciences*. **11**, 535–543 (2007).

4. P. Belin, S. Fecteau, C. Bédard, Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*. **8**, 129–135 (2004).

5. F. Neuner, S. R. Schweinberger, Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain Cogn*. **44**, 342–366 (2000).

6. A. W. Young, V. Bruce, Understanding person perception. *Br J Psychol*. **102**, 959–974 (2011).

7. A. W. Young, D. C. Hay, A. W. Ellis, The faces that launched a thousand slips: everyday difficulties and errors in recognizing people. *Br J Psychol*. **76 (Pt 4)**, 495–523 (1985).

8. D. Y. Tsao, S. Moeller, W. A. Freiwald, Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences*. **105**, 19514–19519 (2008).

9. C. Fisher, W. A. Freiwald, Contrasting specializations for facial motion within the macaque face-processing system. *Curr. Biol*. **25**, 261–266 (2015).

10. M. A. Pinsk *et al*., Neural Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study. *Journal of Neurophysiology*. **101**, 2581–2600 (2009).

11. D. Y. Tsao, N. Schweers, S. Moeller, W. A. Freiwald, Patches of face-selective cortex in the macaque frontal lobe. *Nat Neurosci*. **11**, 877–879 (2008).

12. C. G. Gross, C. E. Rocha-Miranda, D. B. Bender, Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*. **35**, 96–111 (1972).

13. N. Kanwisher, J. McDermott, M. M. Chun, The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).

14. E. B. Issa, J. J. DiCarlo, Precedence of the eye region in neural processing of faces. *Journal of Neuroscience*. **32**, 16666–16682 (2012).

15. D. B. T. McMahon, A. P. Jones, I. V. Bondar, D. A. Leopold, Face-selective neurons maintain consistent visual responses across months. *Proceedings of the National Academy of Sciences*. **111**, 8251–8256 (2014).

16. P. L. Aparicio, E. B. Issa, J. J. DiCarlo, Neurophysiological Organization of the Middle Face Patch in Macaque Inferior Temporal Cortex. *Journal of Neuroscience*. **36**, 12729–12745 (2016).

17. L. Chang, D. Y. Tsao, The Code for Facial Identity in the Primate Brain. *Cell*. **169**, 1013–1028.e14 (2017).

18. W. A. Freiwald, D. Y. Tsao, Functional Compartmentalization and Viewpoint Generalization Within the Macaque Face-Processing System. *Science*. **330**, 845–851 (2010).

19. W. A. Freiwald, D. Y. Tsao, M. S. Livingstone, A face feature space in the macaque temporal lobe. *Nat Neurosci*. **12**, 1187–1196 (2009).

20. S. Ohayon, W. A. Freiwald, D. Y. Tsao, What makes a cell face selective? The importance of contrast. *Neuron*. **74**, 567–581 (2012).

21. J. Taubert, G. Van Belle, W. Vanduffel, B. Rossion, R. Vogels, Neural Correlate of the Thatcher Face Illusion in a Monkey Face-Selective Patch. *J. Neurosci.* **35**, 9872–9878 (2015).

22. J. Taubert, G. Van Belle, W. Vanduffel, B. Rossion, R. Vogels, The effect of face inversion for neurons inside and outside fMRI-defined face-selective cortical regions. *Journal of Neurophysiology*. **113**, 1644–1655 (2015).

23. D. Y. Tsao, W. A. Freiwald, R. B. H. Tootell, M. S. Livingstone, A cortical region consisting entirely of face-selective cells. *Science*. **311**, 670–674 (2006).

24. S. Moeller, W. A. Freiwald, D. Y. Tsao, Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science*. **320**, 1355–1359 (2008).

25. M. I. Gobbini, J. V. Haxby, Neural systems for recognition of familiar faces. *Neuropsychologia*. **45**, 32–41 (2007).

26. L. R. Squire, C. E. L. Stark, R. E. Clark, The medial temporal lobe. *Annu. Rev. Neurosci*. **27**, 279–306 (2004).

27. Y. Miyashita, H. S. Chang, Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*. **331**, 68–70 (1988).

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

29. C. Bruce, R. Desimone, Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology* (1981).

30. D. I. Perrett, E. T. Rolls, W. Caan, Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res*. **47**, 329–342 (1982).

31. R. Desimone, T. D. Albright, C. G. Gross, C. Bruce, Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci*. **4**, 2051–2062 (1984).

32. G. C. Baylis, E. T. Rolls, C. M. Leonard, Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res*. **342**, 91–102 (1985).

33. M. E. Hasselmo, E. T. Rolls, G. C. Baylis, The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*. **32**, 203–218 (1989).

34. E. T. Rolls, M. J. Tovee, Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*. **73**, 713–726 (1995).

35. N. K. Logothetis, H. Guggenberger, S. Peled, J. Pauls, Functional imaging of the monkey brain. *Nat Neurosci*. **2**, 555–562 (1999).

36. W. Vanduffel, Q. Zhu, G. A. Orban, Monkey cortex through fMRI glasses. *Neuron*. **83**, 533–550 (2014).

37. D. Y. Tsao, W. A. Freiwald, T. A. Knutsen, J. B. Mandeville, R. B. H. Tootell, Faces and objects in macaque cerebral cortex. *Nat Neurosci*. **6**, 989–995 (2003).

38. P. Grimaldi, K. S. Saleem, D. Tsao, Anatomical Connections of the Functionally Defined "Face Patches" in the Macaque Monkey. *Neuron*. **90**, 1325–1342 (2016).

39. D. I. Perrett, J. K. Hietanen, M. W. Oram, P. J. Benson, Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond., B, Biol. Sci*. **335**, 23–30 (1992).

40. E. T. Rolls, H. D. Critchley, A. S. Browning, K. Inoue, Face-selective and auditory neurons in the primate orbitofrontal cortex. *Exp Brain Res*. **170**, 74–87 (2006).

41. F. Hadj-Bouziane, A. H. Bell, T. A. Knusten, L. G. Ungerleider, R. B. H. Tootell, Perception of emotional expressions is independent of face selectivity in monkey inferior temporal cortex. *Proceedings of the National Academy of Sciences*. **105**, 5591–5596 (2008).

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

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

44. R. A. Johnston, A. J. Edmonds, Familiar and unfamiliar face recognition: A review. *Memory*. **17**, 577–596 (2009).

45. A. W. Young, A. M. Burton, Are We Face Experts? *Trends in Cognitive Sciences*. **22**, 100–110 (2018).

46. R. Jenkins, A. M. Burton, Stable face representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **366**, 1671–1683 (2011).

47. L. Schwartz, G. Yovel, The roles of perceptual and conceptual information in face recognition. *J Exp Psychol Gen*. **145**, 1493–1511 (2016).

48. J. Haxby, E. Hoffman, M. Gobbini, The distributed human neural system for face perception. *Trends in Cognitive Sciences*. **4**, 223–233 (2000).

49. V. Natu, A. J. O'Toole, The neural processing of familiar and unfamiliar faces: a review and synopsis. *Br J Psychol*. **102**, 726–747 (2011).

50. H. Blank, N. Wieland, K. von Kriegstein, Person recognition and the brain: merging evidence from patients and healthy individuals. *Neurosci Biobehav Rev*. **47**, 717–734 (2014).

51. G. Pourtois, S. Schwartz, M. L. Seghier, F. Lazeyras, P. Vuilleumier, View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: an event-related fMRI study. *NeuroImage*. **24**, 1214–1224 (2005).

52. J. S. Guntupalli, K. G. Wheeler, M. I. Gobbini, Disentangling the Representation of Identity from Head View Along the Human Face Processing Pathway. *Cereb. Cortex*. **27**, 46–53 (2017).

53. N. Kriegeskorte, E. Formisano, B. Sorger, R. Goebel, Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*. **104**, 20600–20605 (2007).

54. R. J. Von Der Heide, Anterior temporal face patches: a meta-analysis and empirical study, 1–18 (2013).

55. L. Garrido *et al*., Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics. *Brain*. **132**, 3443–3455 (2009).

56. M. Mur, D. A. Ruff, J. Bodurka, P. A. Bandettini, N. Kriegeskorte, Face-Identity Change Activation Outside the Face System: "Release from Adaptation" May Not Always Indicate Neuronal Selectivity. *Cerebral Cortex*. **20**, bhp272–2042 (2010).

57. V. Axelrod, G. Yovel, The challenge of localizing the anterior temporal face area: a possible solution. *NeuroImage*. **81**, 371–380 (2013).

58. J. Sliwa, J.-R. Duhamel, O. Pascalis, S. Wirth, Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences*. **108**, 1735–1740 (2011).

59. S. Eifuku, Neuronal Correlates of Face Identification in the Monkey Anterior Temporal Cortical Areas. *Journal of Neurophysiology*. **91**, 358–371 (2003).

60. M. P. Young, S. Yamane, Sparse population coding of faces in the inferotemporal cortex. *Science*. **256**, 1327–1331 (1992).

61. S. Eifuku, W. C. De Souza, R. Nakata, T. Ono, R. Tamura, Neural Representations of Personally Familiar and Unfamiliar Faces in the Anterior Inferior Temporal Cortex of Monkeys. *PLoS ONE*. **6**, e18913 (2011).

62. J. Sliwa, A. Planté, J.-R. Duhamel, S. Wirth, Independent Neuronal Representation of Facial and Vocal Identity in the Monkey Hippocampus and Inferotemporal Cortex. *Cerebral Cortex*. **26**, bhu257–966 (2014).

63. J. Munuera, M. Rigotti, C. D. Salzman, Shared neural coding for social hierarchy and reward value in primate amygdala. *Nat Neurosci*. **21**, 415–423 (2018).

64. K. Nakamura, K. Matsumoto, A. Mikami, K. Kubota, Visual response properties of single neurons in the temporal pole of behaving monkeys. *Journal of Neurophysiology*. **71**, 1206–1221 (1994).

65. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing on JSTOR. *Journal of the royal statistical society Series B (* … (1995), doi:10.2307/2346101.

66. J. A. Collins, I. R. Olson, Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*. **61**, 65–79 (2014).

67. V. S. Natu, A. J. O'Toole, Spatiotemporal changes in neural response patterns to faces varying in visual familiarity. *NeuroImage*. **108**, 151–159 (2015).

68. C.-C. Carbon, Famous faces as icons. The illusion of being an expert in the recognition of famous faces. *Perception*. **37**, 801–806 (2008).

69. K. M. Gothard, C. A. Erickson, D. G. Amaral, How do rhesus monkeys (Macaca mulatta) scan faces in a visual paired comparison task? *Anim Cogn*. **7**, 25–36 (2004).

70. K. M. Gothard, K. N. Brooks, M. A. Peterson, Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim Cogn*. **12**, 155–167 (2009).

71. L. A. Parr, J. T. Winslow, W. D. Hopkins, F. B. M. de Waal, Recognizing facial cues: Individual discrimination by chimpanzees (Pan troglodytes) and rhesus monkeys (Macaca mulatta). *Journal of Comparative Psychology*. **114**, 47–60 (2000).

72. C. D. Dahl, N. K. Logothetis, K. L. Hoffman, Individuation and holistic processing of faces in rhesus monkeys. *Proceedings of the Royal Society of London B: Biological Sciences*. **274**, 2069–2076 (2007).

73. I. Adachi, R. R. Hampton, Rhesus monkeys see who they hear: spontaneous cross-modal memory for familiar conspecifics. *PLoS ONE*. **6**, e23345 (2011).

74. C. A. Conway, B. C. Jones, L. M. DeBruine, A. C. Little, A. Sahraie, Transient pupil constrictions to faces are sensitive to orientation and species. *Journal of Vision*. **8**, 17.1–11 (2008).

75. R. B. Ebitz, J. M. Pearson, M. L. Platt, Pupil size and social vigilance in rhesus macaques. *Front Neurosci*. **8**, 100 (2014).

76. C. Fisher, W. A. Freiwald, Whole-agent selectivity within the macaque face-processing system. *Proceedings of the National Academy of Sciences*. **112**, 14717–14722 (2015).

77. W. A. Suzuki, D. G. Amaral, Perirhinal and parahippocampal cortices of the macaque monkey: cytoarchitectonic and chemoarchitectonic organization. *J. Comp. Neurol.* **463**, 67–91 (2003).

78. H. Kondo, K. S. Saleem, J. L. Price, Differential connections of the temporal pole with the orbital and medial prefrontal networks in macaque monkeys. *J. Comp. Neurol.* **465**, 499–523 (2003).

79. D. J. Kravitz, K. S. Saleem, C. I. Baker, L. G. Ungerleider, M. Mishkin, The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*. **17**, 26–49 (2013).

80. I. D. Popivanov, J. Jastorff, W. Vanduffel, R. Vogels, Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *NeuroImage*. **63**, 723–741 (2012).

81. K. S. Saleem, N. K. Logothetis, *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates* (San Diego: Elsevier; Academic Press, 2012).

82. A. M. Burton, S. Wilson, M. Cowan, V. Bruce, Face Recognition in Poor-Quality Video: Evidence From Security Surveillance. *Psychological Science*. **10**, 243–248 (1999).

83. P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. *Proc. IEEE*. **94**, 1948–1962 (2006).

84. M. Ramon, L. Vizioli, J. Liu-Shuang, B. Rossion, Neural microgenesis of personally familiar face recognition. *Proceedings of the National Academy of Sciences*. **112**, E4835–44 (2015).

85. M. Bar, Visual objects in context. *Nat Rev Neurosci*. **5**, 617–629 (2004).

86. K. I. Naka, W. A. Rushton, S-potentials from colour units in the retina of fish (Cyprinidae). *The Journal of Physiology*. **185**, 536–555 (1966).

87. S. M. Landi, W. A. Freiwald, Two areas for familiar face recognition in the primate brain. *Science*. **357**, 591–595 (2017).

88. K. Srihasam, J. L. Vincent, M. S. Livingstone, Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat Neurosci*. **17**, 1776–1783 (2014).

89. M. J. Arcaro, P. F. Schade, J. L. Vincent, C. R. Ponce, M. S. Livingstone, Seeing faces is necessary for face-domain formation. *Nat Neurosci*. **20**, 1404–1412 (2017).

90. G. Golarai, A. Liberman, K. Grill-Spector, Experience Shapes the Development of Neural Substrates of Face Processing in Human Ventral Temporal Cortex. *Cereb. Cortex*. **27**, 1229–1244 (2017).

91. A. Poremba *et al.*, Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature*. **427**, 448–451 (2004).

92. C. Perrodin, C. Kayser, T. J. Abel, N. K. Logothetis, C. I. Petkov, Who is That? Brain Networks and Mechanisms for Identifying Individuals. *Trends in Cognitive Sciences*. **19**, 783–796 (2015).

93. J. Z. Leibo, Q. Liao, F. Anselmi, W. A. Freiwald, T. Poggio, View-Tolerant Face Recognition and Hebbian Learning Imply Mirror-Symmetric Neural Tuning to Head Orientation. *Curr. Biol.* **27**, 62–67 (2017).

94. A. M. Burton, S. R. Schweinberger, R. Jenkins, J. M. Kaufmann, Arguments Against a Configural Processing Account of Familiar Face Recognition. *Perspect Psychol Sci*. **10**, 482–496 (2015).

95. A. Sandford, A. M. Burton, Tolerance for distorted faces: challenges to a configural processing account of familiar face recognition. *Cognition*. **132**, 262–268 (2014).

96. G. J. Hole, P. A. George, K. Eaves, A. Rasek, Effects of geometric distortions on face-recognition performance. *Perception*. **31**, 1221–1240 (2002).

97. M. Bindemann, A. M. Burton, H. Leuthold, S. R. Schweinberger, Brain potential correlates of face recognition: geometric distortions and the N250r brain response to stimulus repetitions. *Psychophysiology*. **45**, 535–544 (2008).

98. V. Bruce *et al*., Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*. **5**, 339–360 (1999).

99. H. D. Ellis, J. W. Shepherd, G. M. Davies, Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*. **8**, 431–439 (1979).

100. A. W. Young, D. C. Hay, K. H. McWeeny, B. M. Flude, A. W. Ellis, Matching familiar and unfamiliar faces on internal and external features. *Perception*. **14**, 737–746 (1985).

101. T. J. Andrews, J. Davies-Thompson, A. Kingstone, A. W. Young, Internal and External Features of the Face Are Represented Holistically in Face-Selective Regions of Visual Cortex. *J. Neurosci*. **30**, 3544–3552 (2010).

102. N. N. Watier, C. A. Collin, Effects of familiarity on spatial frequency thresholds for face matching. *Perception*. **38**, 1497–1507 (2009).

103. G. R. Loftus, E. M. Harley, Why is it easier to identify someone close than far away? *Psychon Bull Rev*. **12**, 43–65 (2005).

104. L. Li, E. K. Miller, R. Desimone, The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*. **69**, 1918–1929 (1993).

105. H. Sawamura, G. A. Orban, R. Vogels, Selectivity of Neuronal Adaptation Does Not Match Response Selectivity: A Single-Cell Study of the fMRI Adaptation Paradigm. *Neuron*. **49**, 307–318 (2006).

106. T. Valentine, A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q J Exp Psychol A*. **43**, 161–204 (1991).

107. M. Lewis, Face-space-R: Towards a unified account of face recognition. *Visual Cognition*. **11**, 29–69 (2004).

108. M. Bernstein, J. Oron, B. Sadeh, G. Yovel, An integrated face-body representation in the fusiform gyrus but not the lateral occipital cortex. *J Cogn Neurosci*. **26**, 2469–2478 (2014).

109. Y. Song, Y. L. L. Luo, X. Li, M. Xu, J. Liu, Representation of contextually related multiple objects in the human ventral visual pathway. *J Cogn Neurosci*. **25**, 1261–1269 (2013).

110. B. Pascual *et al*., Large-scale brain networks of the human left temporal pole: a functional connectivity MRI study. *Cereb. Cortex*. **25**, 680–702 (2015).

111. L. Fan *et al*., Connectivity-based parcellation of the human temporal pole using diffusion tensor imaging. *Cereb. Cortex*. **24**, 3365–3378 (2014).

112. C. I. Petkov *et al*., Different forms of effective connectivity in primate frontotemporal pathways. *Nat Commun*. **6**, 6000 (2015).

113. R. Desimone, C. G. Gross, Visual areas in the temporal cortex of the macaque. *Brain Res*. **178**, 363–380 (1979).

114. D. J. Freedman, M. Riesenhuber, T. Poggio, E. K. Miller, Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*. **16**, 1631–1644 (2006).

115. C. I. Baker, M. Behrmann, C. R. Olson, Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat Neurosci*. **5**, 1210–1216 (2002).

116. E. Kobatake, G. Wang, K. Tanaka, Effects of Shape-Discrimination Training on the Selectivity of Inferotemporal Cells in Adult Monkeys. *Journal of Neurophysiology*. **80**, 324–330 (1998).

117. M. C. Booth, E. T. Rolls, View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*. **8**, 510–523 (1998).

118. S.-P. Ku, A. S. Tolias, N. K. Logothetis, J. Goense, fMRI of the Face-Processing Network in the Ventral Temporal Lobe of Awake and Anesthetized Macaques. *Neuron*. **70**, 352–362 (2011).

119. G. Wang, K. Tanaka, M. Tanifuji, Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*. **272**, 1665–1668 (1996).

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

121. A. H. Bell *et al*., Relationship between Functional Magnetic Resonance Imaging-Identified Regions and Neuronal Category Selectivity. *J. Neurosci*. **31**, 12229–12240 (2011).

122. N. Miyakawa *et al*., Heterogeneous Redistribution of Facial Subcategory Information Within and Outside the Face-Selective Domain in Primate Inferior Temporal Cortex. *Cerebral Cortex*. **28**, 1416–1431 (2018).

123. W. Dittrich, How monkeys see others: Discrimination and recognition of monkeys' shape. *Behav. Processes*. **33**, 139–154 (1994).

124. T. Matsuno, K. Fujita, Body inversion effect in monkeys. *PLoS ONE*. **13**, e0204353 (2018).

125. L. Chang, Q. Fang, S. Zhang, M.-M. Poo, N. Gong, Mirror-induced self-directed behaviors in rhesus monkeys after visual-somatosensory training. *Curr. Biol*. **25**, 212–217 (2015).

126. R. A. Robbins, M. Coltheart, The effects of inversion and familiarity on face versus body cues to person recognition. *Journal of Experimental Psychology: Human Perception and Performance*. **38**, 1098–1104 (2012).

127. D. Kaiser, L. Strnad, K. N. Seidl, S. Kastner, M. V. Peelen, Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *Journal of Neurophysiology*. **111**, 82–90 (2014).

128. C. A. Hahn, A. J. O'Toole, P. J. Phillips, Dissecting the time course of person recognition in natural viewing environments. *British Journal of Psychology*. **107**, 117–134 (2016).

129. D. Rendall, P. S. Rodman, R. E. Emond, Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour*. **51**, 1007–1015 (1996).

130. H. M. Habbershon, S. Z. Ahmed, Y. E. Cohen, Rhesus macaques recognize unique multimodal face-voice relations of familiar individuals and not of unfamiliar ones. *Brain* (2013).

131. J. R. Hanley, S. T. Smith, J. Hadfield, I Recognise you but I Can't Place you: An Investigation of Familiar-only Experiences during Tests of Voice and Face Recognition. *Q J Exp Psychol A*. **51**, 179–195 (1998).

132. J. R. Hanley, J. M. Turner, Why are familiar-only experiences more frequent for voices than for faces? *Q J Exp Psychol A*. **53**, 1105–1116 (2000).

133. L. Damjanovic, J. R. Hanley, Recalling episodic and semantic information about famous faces and voices. *Mem Cognit*. **35**, 1205–1210 (2007).

134. C. Perrodin, C. Kayser, N. K. Logothetis, C. I. Petkov, Voice cells in the primate temporal lobe. *Curr. Biol*. **21**, 1408–1415 (2011).

135. Y. Kikuchi, B. Horwitz, M. Mishkin, Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *Journal of Neuroscience*. **30**, 13021–13030 (2010).

136. C. Perrodin, C. Kayser, N. K. Logothetis, C. I. Petkov, Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *Journal of Neuroscience*. **34**, 2524–2537 (2014).

137. H. E. Moss, J. M. Rodd, E. A. Stamatakis, P. Bright, L. K. Tyler, Anteromedial temporal cortex supports fine-grained differentiation among objects. *Cerebral Cortex*. **15**, 616–627 (2005).

138. L. K. Tyler *et al*., Objects and categories: feature statistics and object processing in the ventral stream. *J Cogn Neurosci*. **25**, 1723–1735 (2013).

139. P. Fries, A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*. **9**, 474–480 (2005).

140. G. G. Gregoriou, S. J. Gotts, H. Zhou, R. Desimone, High-Frequency, Long-Range Coupling Between Prefrontal and Visual Cortex During Attention. *Science*. **324**, 1207–1210 (2009).

141. B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, R. A. Andersen, Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat Neurosci*. **5**, 805–811 (2002).

142. A. Afraz, E. S. Boyden, J. J. DiCarlo, 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 (2015).

143. S. Sadagopan, W. Zarco, W. A. Freiwald, A causal relationship between face-patch activity and face-detection behavior. *Elife*. **6**, 6730 (2017).

144. V. Willenbockel *et al*., Controlling low-level image properties: the SHINE toolbox. *Behav Res Methods*. **42**, 671–684 (2010).

145. T. J. Barry, J. W. Griffith, S. De Rossi, D. Hermans, Meet the Fribbles: novel stimuli for use within behavioural research. *Front Psychol*. **5**, 103 (2014).

146. W. G. Hayward, M. J. Tarr, Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*. **23**, 1511–1521 (1997).

147. J. E. Hummel, I. Biederman, Dynamic binding in a neural network for shape recognition. *Psychol Rev*. **99**, 480–517 (1992).

148. T. Meyer, C. Walker, R. Y. Cho, C. R. Olson, Image familiarization sharpens response dynamics of neurons in inferotemporal cortex. *Nat Neurosci*. **17**, 1388–1394 (2014).

149. A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*. **9**, 179–194 (1999).

150. B. Fischl, M. I. Sereno, A. M. Dale, Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*. **9**, 195–207 (1999).

151. S. Frey *et al*., An MRI based average macaque monkey stereotaxic atlas and space (MNI monkey space). *NeuroImage*. **55**, 1435–1442 (2011).

152. G. Paxinos, X. F. Huang, A. W. Toga, The Rhesus Monkey Brain in Stereotaxic Coordinates (2000).

153. P. Polosecki *et al*., Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. *Journal of Neuroscience*. **33**, 11768–11773 (2013).

154. K. Marciniak, A. Atabaki, P. W. Dicke, P. Thier, Disparate substrates for head gaze following and face perception in the monkey superior temporal sulcus. *Elife*. **3**, 161 (2014).

155. D. G. Leibovici, S. Smith, Comparing groups of subjects in fMRI studies: a review of the GLM approach (Oxford, UK). Available at: http://www.fmrib.ox.ac.uk/analysis/techrep/tr00dl1/tr00dl1/index.html. *Oxford) Available at www fmrib ox ac uk …* (2000).

156. J. Jastorff, I. D. Popivanov, R. Vogels, W. Vanduffel, G. A. Orban, Integration of shape and motion cues in biological motion processing in the monkey STS. *NeuroImage*. **60**, 911–921 (2012).

157. M. A. Pinsk *et al*., Neural Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study. *Journal of Neurophysiology*. **101**, 2581–2600 (2008).

158. N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, C. I. Baker, Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*. **12**, 535–540 (2009).

159. S. Holm, A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), doi:10.2307/4615733.

160. G. Loffler, G. Yourganov, F. Wilkinson, H. R. Wilson, fMRI evidence for the neural representation of faces. *Nat Neurosci*. **8**, 1386–1390 (2005).

161. B. Efron, R. J. Tibshirani, *An introduction to the bootstrap* (Chapman and Hall, 1993).