

The Macaque Face Patch System: A Window into Object Representation

DORIS TSAO

Division of Biology and Biological Engineering and Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125

Correspondence: dortsao@caltech.edu

The macaque brain contains a set of regions that show stronger fMRI activation to faces than other classes of object. This “face patch system” has provided a unique opportunity to gain insight into the organizing principles of IT cortex and to dissect the neural mechanisms underlying form perception, because the system is specialized to process one class of complex forms, and because its computational components are spatially segregated. Over the past 5 years, we have set out to exploit this system to clarify the nature of object representation in the brain through a multilevel approach combining electrophysiology, anatomy, and behavior. These experiments reveal (1) a remarkably precise connectivity of face patches to each other, (2) a functional hierarchy for representation of view-invariant identity comprising at least three distinct stages along the face patch system, and (3) the computational mechanisms used by cells in face patches to detect and recognize faces, including measurement of diagnostic local contrast features for detection and measurement of face feature values for recognition.

How does the brain represent objects? This question had its beginnings in philosophy. Our fundamental intuition of the physical world consists of a space containing objects, and philosophers starting from Plato wondered about the basis for the percept of these “pure forms” (e.g., *the tree*) that were clearly different from any real instance. Very early on, the mind could already sense something mysterious about the problem of object perception. Object representation constitutes the basic infrastructure on which the brain operates. We speak in nouns; we remember people, places, and things; and we think in terms of concepts, which can be construed as a generalization of objects. Despite its clear importance, we still understand very little about the neural basis for object perception. In particular, to understand visual object perception, three critical problems need to be solved: (1) How is an object first generated (i.e., how are retinal pixels stitched together into units)? (2) How are these stitched units identified? (3) How are identified units relayed to higher-order brain areas to enable flexible behavior? During the past decade, work in my laboratory has focused largely on the second question, addressing the mechanisms for face processing in macaque inferotemporal (IT) cortex. Here, I describe what we have learned about principles of object identification in the brain from studying a set of regions in the temporal lobe specialized for face processing, the macaque face patch system.

The first stage of visual information processing in the cortex occurs in area V1, where cells extract local stimulus properties like edge orientation, motion, and color contrast. Then, visual information is transmitted through a series of additional stages, V2, V3, V4, which each contain a retinotopic map of space, and must be performing local computations beyond edge detection. The precise

nature of these steps remains a mystery. One major transformation appears to be segmentation (i.e., organizing visual information into discrete pieces corresponding to different objects) (Zhou et al. 2000; Bushnell et al. 2011), a highly challenging task owing to partial occlusion and the need to interpolate illusory contours (Fig. 1). Then, visual information proceeds to a large brain region called inferotemporal (IT) cortex, which has been strongly implicated in high-level object recognition (e.g., recognizing a rose, a bird, or a face). A lesion to this part of the brain can create an inability to recognize specific classes of objects such as faces, suggesting this is an important brain area to study if we want to understand object perception.

How are objects represented in IT cortex? Charles Gross and coworkers reported discovery of cells in the temporal lobe that were selective for complex forms such as hands, trees, and faces (Bruce et al. 1981), but the difficulty of finding these cells precluded deeper understanding. In 1997, Nancy Kanwisher, using fMRI in humans, reported the discovery of a face-selective area in the brain (Kanwisher et al. 1997). Remarkably, this area seemed to be in the same place in every subject she scanned, suggesting that face processing occurs in a discrete chunk of cortex. Although this finding was provocative and exciting, it remained a mystery what the cells in these regions might be doing, as the region was found using fMRI, and the relationship between blood flow measured by fMRI and underlying neural activity remains an area of active research (Logothetis 2008; Schummers et al. 2008). Most importantly, fMRI measures activity at a spatial scale of $\sim 1 \text{ mm}^3$, whereas neural activity is organized at a much finer scale such that even neighboring cells can have very different tuning properties (Ohki et al. 2006).

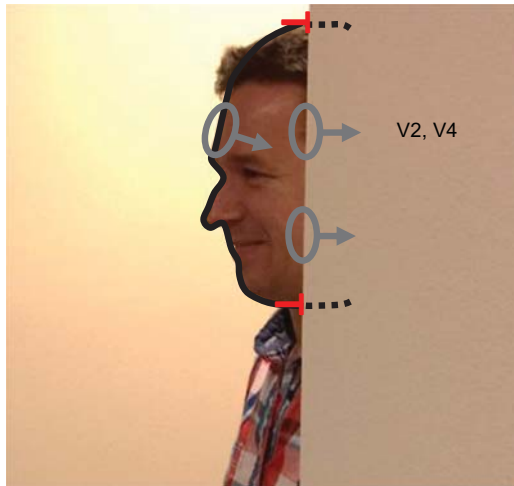


Figure 1. Segmentation processes in extrastriate retinotopic cortex. In area V4, a “boundary curvature” cell tuned for a right angle at south would be suppressed, because the presence of the T junction (red) would signal to the cell that the boundary continues behind the wall instead of making a 90° turn (Bushnell et al. 2011). In area V2, cells tuned for border-ownership (gray) configure their activity to signal the correct ownership of all the contours in the image (Zhou et al. 2000). In this way, a map is generated not just of the location of edges in the image, but which figure owns them, and how they continue behind occluders.

To clarify the link between face cells and fMRI-identified face areas, we performed fMRI experiments in alert monkeys. We found not just one such area, but six of them (Fig. 2A). Moreover, the fact that these six “face patches” were located in the same place across the two hemispheres, and in similar locations across animals, gave the first hint that they constitute a system and not just random islands of face-selective cortex. To study the selectivity of single neurons in these patches, we targeted electrodes to ML/MF, AL, and AM, and asked what the responses of cells in these regions were to the same stimuli that we used in the fMRI localizer experiment. We found that all three regions contained a very high percentage of face-selective cells, with 97% of visually responsive cells in ML/MF giving a mean response to faces at least twice as strong as to other objects (Tsao et al. 2006). This finding was exciting because it meant we now had a system where we could systematically dissect how one visual form is represented.

The macaque face patch system has provided a unique opportunity to gain insight into the organizing principles of IT cortex and to dissect the neural mechanisms underlying form perception, because the system is specialized to process one class of complex forms, and because its computational components are spatially segregated. Over

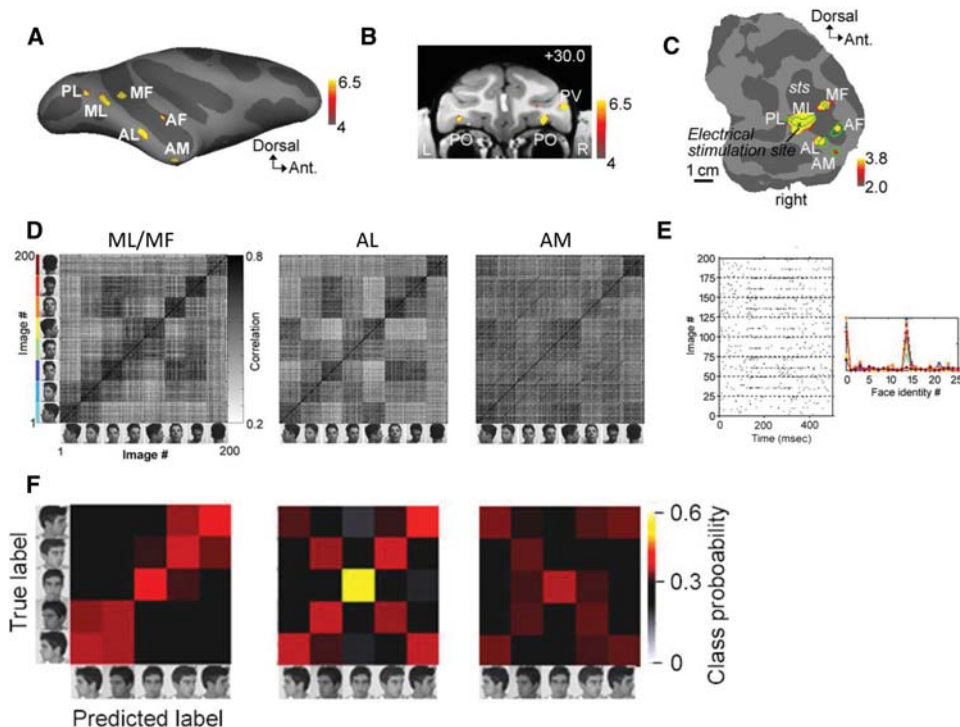


Figure 2. Dissecting face processing in the monkey. (A) Six face patches shown on inflated right hemisphere of macaque brain (Tsao et al. 2008a). (B) Two prefrontal face-selective patches, PO in the lateral orbital sulcus and PV in the infrapincipal dimple (Tsao et al. 2008b). (C) Connectivity of temporal face patches revealed by microstimulation targeted to face patch ML combined with fMRI; areas significantly activated by microstimulation overlaid on a flatmap (Moeller et al. 2008). (D) Population similarity matrices in the three face patches. A 200×200 matrix of correlation coefficients was computed between responses of all visually responsive cells to a set of 200 stimuli (consisting of 25 different identities each at eight different head orientations) from ML/MF ($N = 121$ cells), AL ($N = 189$ cells), and AM ($N = 158$ cells) (Freiwald and Tsao 2010). (E) Mean response time courses of an example sparse, view-invariant identity selective cell from AM to the 200 stimuli. (Right) Mean response levels to the 25 individuals at each head orientation (Freiwald and Tsao 2010). (F) Decoding of view from fMRI responses in ML/MF, AL, and AM to four identities each at five views (Dubois et al. 2015).

the past 5 years, my laboratory has set out to exploit this system to clarify the nature of object representation in the brain through a multilevel approach combining electrophysiology, anatomy, and behavior, focusing on three questions.

- *Connectivity*: What is the anatomical wiring diagram of the face patches?
- *Functional architecture*: Are the six patches performing different functions?
- *Computational mechanisms*: What are the mechanisms for detection and recognition of faces used by cells in the face patches?

CONNECTIVITY

The existence of six face patches raised obvious questions about anatomical connectivity. Do the patches form a unified system, or is each patch processing faces independently of the others? Does the anatomy reveal any hierarchical relationships? What are the downstream outputs of face patches? To image connectivity of the face patches *in vivo*, we electrically microstimulated different face patches while the monkey was inside the fMRI scanner (Moeller et al. 2008). Whenever we stimulated one patch, the other patches would light up, but not the surrounding cortex, indicating the patches are strongly connected to each other but not to the other parts of IT cortex (Fig. 2C). In addition, stimulation of face patches activated specific subregions of three subcortical areas: the amygdala, claustrum, and pulvinar. More recently, we have confirmed these results with fMRI-guided anatomical tracer injections (Grimaldi et al. 2012, 2013).

FUNCTIONAL ARCHITECTURE

Because the six face patches span the entire extent of the temporal lobe, it seemed likely that each patch performs a unique function. To discover functional differences between patches, we presented several large sets of face stimuli to animals while recording from multiple patches. In one of these experiments, we presented 25 different identities each at eight different head orientations and discovered that a major functional distinction between the patches concerns how they represent identity across different views (Freiwald and Tsao 2010). Neurons in ML and MF are view-specific; neurons in AL are tuned to identity mirror-symmetrically across views, thus achieving partial view invariance; and neurons in AM, the most anterior face patch, achieve almost full view invariance (Fig. 2D). We further discovered a remarkable cell type in the most anterior face patch AM, which responds extremely sparsely to only a small subset of face identities, invariantly across changes in view (Fig. 2E). Thus, it appears a major goal of the face patches is to build, in stepwise fashion, a representation of individual identity invariant to view direction.

Is there any spatial organization to view and identity tuning? To address this, we presented four identities each

at five head orientations in a block-design fMRI experiment (a subset of the stimuli used in Freiwald and Tsao 2010) and found that multivoxel pattern analysis on the fMRI responses from ML/MF, AL, and AM could successfully decode view. Moreover, the view decoding made mirror-symmetric mistakes in AL and AM, just as we had found earlier in the units (Fig. 2F) (Dubois et al. 2015). This suggests that cells tuned to the same view are spatially clustered in each face patch.

Comparison of stimulus selectivity across different patches has revealed other significant differences. There is a clear change in species selectivity going from ML/MF, where most cells respond vigorously to both monkey and human faces, to AM, where many cells are selective for either monkey or human faces (Moeller and Tsao 2011). Experiments in which we presented random face fragments revealed that the effective fragments of a face that trigger firing increase in size and complexity going from posterior to anterior face patches (Cheng et al. 2013). Underscoring this progression in size and complexity, in the most posterior patch PL, Issa and DiCarlo (2012) found that the most effective fragment was the contralateral eye. Overall, our experiments indicate a sparser, more holistic, and more invariant representation as one proceeds anteriorly along the face patch system, consistent with the finding of “Jennifer Aniston cells” one step further in the medial temporal lobe (Quiroga et al. 2005). We do not yet understand the fundamental principle governing why each patch processes faces only up to a certain level of complexity before handing the problem off to the next patch; it seems clear that a deep answer to this question would require not just documentation of phenomenological differences between patches, but a grasp of the fundamental computational architecture.

COMPUTATIONAL MECHANISMS

As a first foray into understanding the computational architecture of the face patches, we have delved into the detailed mechanisms used by single cells to detect and recognize faces, exploiting easily-to-parameterize cartoon faces.

The first step in face processing is face detection (i.e., detecting a face is present somewhere regardless of whose face it is). Faces are robustly detected by computer vision algorithms that search for characteristic coarse contrast features (Viola and Jones 2001; Sinha 2002) (e.g., eyes darker than nose). If one examines the contrast between pairs of regions when a face is illuminated under a large variety of conditions, one finds that for some features, such as upper lip and cheek, there is no consistent contrast relationship. The upper lip is sometimes darker and sometimes brighter than the cheek, depending on the lighting. But for other features, there is a consistent contrast relationship (e.g., the nose is always brighter than the left eye).

Pawan Sinha (2002) suggested that for face detection, the most important features should be ones that are invariant to changes in lighting. To test whether cells in the face patches might be using these illumination-invariant

contrast features to detect faces, we constructed an artificial face stimulus consisting of 11 different regions varying in brightness. Individual cells from the middle face patch showed a wide range of responsiveness to these part intensity stimuli, with some stimuli eliciting stronger responses than to a real face, and others eliciting no response at all (Fig. 3A). To determine whether contrast between pairs of parts might be driving this variation, for each pair of parts, we computed the mean response when part A was darker than part B and the mean response when part A was brighter than part B for each of the 55 pairs of parts. Figure 3B shows results for an example unit, and Figure 3C shows the result for the whole population. Remarkably, the cells were completely consistent in their contrast preference (e.g., almost 100 cells preferred the left eye to be darker than the nose, and not a single cell preferred the opposite contrast relationship). Moreover, the preferred features were completely consistent with those predicted from the light-invariance experiments, indicated by the purple arrows. One question often asked about face cells is how do we know these cells are really coding faces and not some other object that we simply have not shown yet, because we obviously cannot show every possible object to a single cell in IT cortex. The consistency of contrast preferences of face cells, with each other and with computational light-invariance experiments, is powerful evidence that these cells are truly coding faces. At the same time, the result shows that these cells are using more primitive mecha-

nisms to detect faces than human observers. Even though both stimuli in Figure 3A appear “face-like” to human observers, they could elicit very different responses in a subset of face cells in ML/MF.

Indeed, the fact that we see both stimuli as faces suggests that contrast cannot be the whole story to face detection. We can readily see faces in line drawings in which there is no contrast. Thus, feature shape must also play an important role. What is the contribution of feature shape to face detection? To address this question, we recorded responses of cells in ML/MF to a cartoon face defined by seven different elementary parts. Responses to the 128 combinations of these seven face parts showed that individual cells are selective for the presence of specific face parts, such as eyes or hair. Figure 4A shows responses of two example cells from the middle face patch to these 128 stimuli, illustrating selectivity for different parts. This result is interesting because it challenges one of the long-standing assumptions about IT cortex—namely, that it is organized into feature columns, like V1, with each column processing various moderately complex shapes that are visually similar (Tanaka 2003). A pair of disks and an upside-down U have nothing visually similar about them. Rather, what they have in common is that they are both defining features of a face, an ethologically meaningful unit. We found neighboring cells within the face patch tuned to such visually dissimilar features, as well as single cells tuned to multiple such features. Thus, the ethological meaning of objects is clearly an important driving

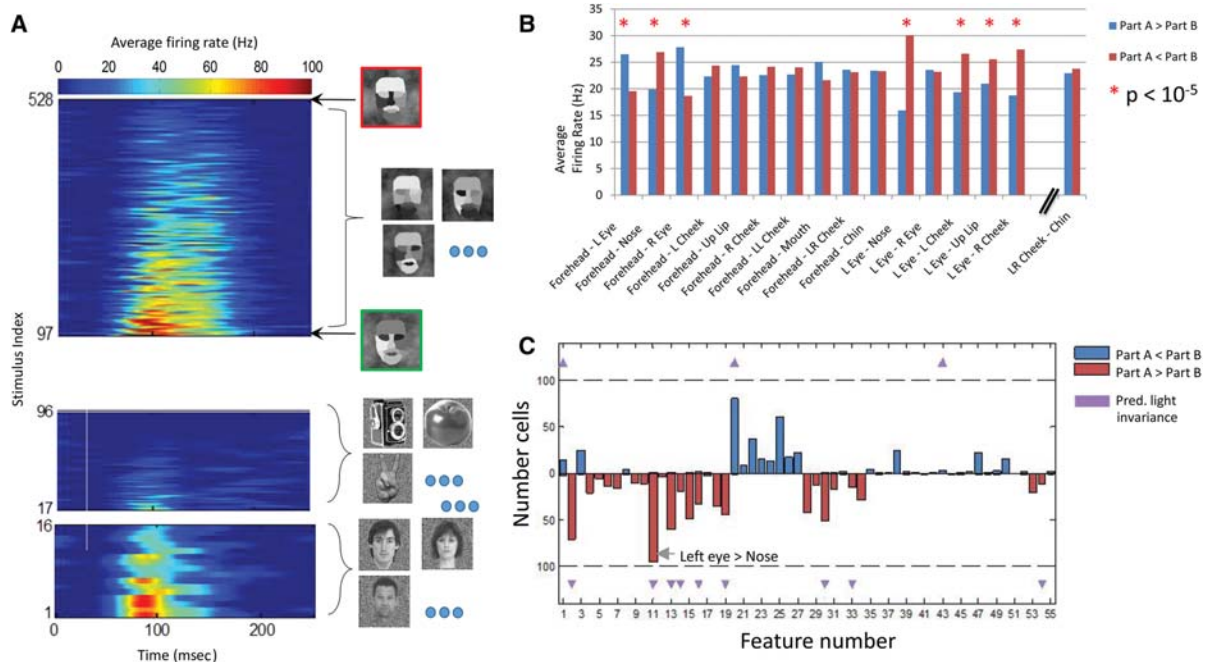


Figure 3. Detecting faces through selectivity for characteristic contrast features (Ohayon et al. 2012). (A) Response of an example cell from ML/MF to 16 pictures of real faces (bottom), 80 pictures of nonface objects (middle), and 432 part intensity stimuli constructed by randomly varying the brightness of 12 face regions. An example ineffective (red outline) and effective (green outline) part intensity stimulus are shown. (B) Responses of an example cell from ML/MF to a subset of the 55 feature pairs, showing mean response to both contrast polarities of each pair. Asterisks mark feature pairs for which the cell showed significant contrast selectivity. (C) Significant contrast feature histogram. Blue (red) bars indicate the number of cells tuned for intensity in A greater (less) than intensity in B. Triangles indicate predictions from computational light-invariance experiments.

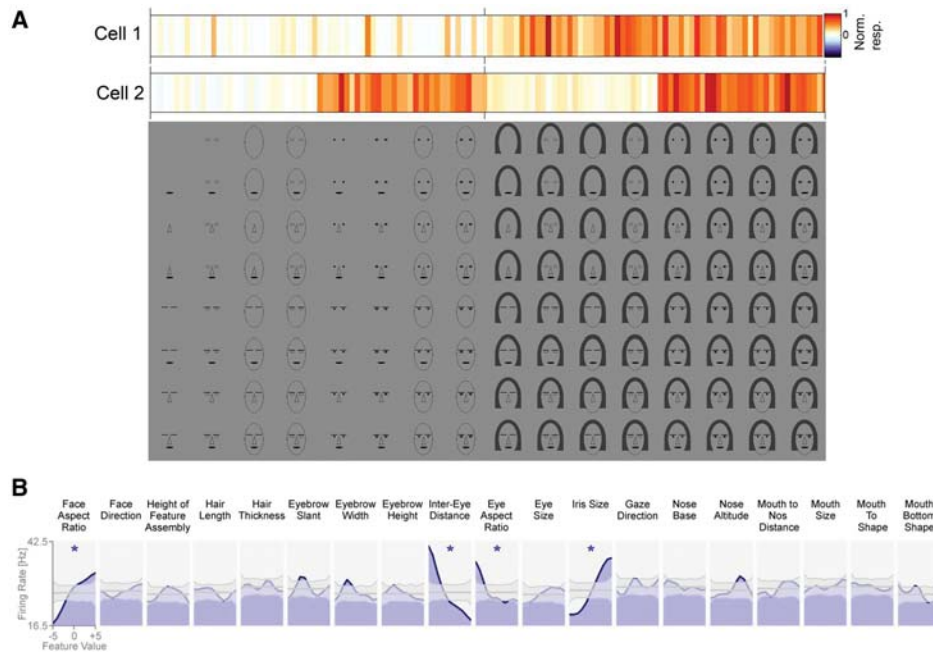


Figure 4. Probing mechanisms for face detection and recognition with cartoon faces. (A) Responses of two examples cells from ML/MF to 128 combinations of seven cartoon face parts. Cell 1 was selective for the presence of hair, cell 2 for the presence of irises. (B) Tuning of an example cell from ML/MF to 19 cartoon face dimensions. Tuning curve significantly deviating from a shuffle control are indicated by asterisk. This cell was tuned to four parameters: face aspect ratio, inter-eye distance, eye aspect ratio, and iris size.

force in IT organization, above and beyond low-level visual feature similarity.

It is critically important for primates to not only detect other faces, but to recognize them individually. What is the neural mechanism for distinguishing different faces? In general terms, this could be accomplished based on the overall shape of the face (e.g., narrow vs. round), the shape of specific features (e.g., iris size), or the spatial relationship between different features (e.g., inter-eye distance). To distinguish these possibilities, we constructed another set of cartoon faces, this time varied in identity. Each cartoon face was defined by 19 dimensions, and the values of the dimensions were varied randomly and independently; some dimensions described the overall shape of the face, some described the shape of specific features, and some described the spatial relationship between features. We found that individual cells are tuned to subsets of face features. Figure 4B shows tuning curves of an example cell to the 19 feature dimensions; this cell was significantly tuned to four features, face aspect ratio, inter-eye distance, eye aspect ratio, and iris size. Interestingly, all four of the tuning curves are ramp shaped, with a maximum at one extreme and a minimum at the opposite extreme. This was true across the population, suggesting that these cells are acting like rulers, which is consistent with a “face space” representation (Valentine et al. 2015), where cells are measuring deviation from the average face along specific axes rather than encoding specific exemplars. This preference for extreme feature values may explain the power of caricatures, which would be stimulating the population to fire at its maximum dynamic range.

Obviously, one limitation with these cartoon experiments is that it is unclear how the principles we have covered generalize to encoding real faces. For example, if one constructs a realistic face space by performing principle components analysis on a large set of real faces, do cells in the face patches also show ramp-shaped tuning to the realistic face dimensions? And how well can one decode identity of real faces from face patch population activity? If cells truly are encoding specific axes through linear ramps, this suggests that a simple linear regression should be sufficient to decode facial identity. We are currently addressing these questions through ongoing experiments.

SUMMARY AND OUTLOOK

The macaque face patch system is a remarkable gift of nature for understanding the steps of object representation. Even though we are only just beginning to understand the principles underlying the organization of this system, it is already clear that major computational transformations are accomplished between each stage, to generate a code for facial identity in the most anterior face patch AM invariant to transformations such as view, position, and size. Future work will need to clarify whether and how the organization of this system generalizes to other object categories; evidence suggests that systems in IT cortex comprising multiple patches are also used to represent scenes (Kornblith et al. 2013), bodies (Popivanov et al. 2012, 2014), and colored objects (Lafer-Sousa and Conway 2013).

I believe the biggest questions about the face patch system concern how the patches communicate with the rest of the brain, including earlier retinotopic cortex and higher-order brain areas that ultimately drive behavior. The face patches are like a wonderfully lit house in the middle of the woods. What is needed now is to follow the trail of bread crumbs from them, both forward and backward, to gain a deeper level of understanding into (1) how an object first arises as a coherent unit and how this coherent unit is transmitted as such from retinotopic to IT cortex, and (2) how the code for object identity, represented by a distributed population of neurons, is routed to downstream areas to enable flexible, goal-directed behavior. It is clear these processes must involve globally organized interactions that we only have the barest inkling of so far. For example, if two faces are present, how does the brain keep track of the identity, location, and actions of each separately? This “binding problem” is one of the abiding mysteries of systems neuroscience. In his book *Rhythms of the Brain*, Buzsáki (2006) vividly evokes the excitement that greeted the prospect of an imminent solution to the binding problem. It would be exciting if research on face processing, starting from sure knowledge of where the label for facial identity is located in the brain, could bring us closer to that day.

ACKNOWLEDGMENTS

I thank Winrich Freiwald, who started this journey into the face patch system side-by-side with me, and members of my laboratory past and present whose creativity, insight, skill, and hard work are a joy to acknowledge.

REFERENCES

- Bruce C, Desimone R, Gross CG. 1981. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* **46**: 369–384.
- Bushnell BN, Harding PJ, Kosai Y, Pasupathy A. 2011. Partial occlusion modulates contour-based shape encoding in primate area V4. *J Neurosci* **31**: 4012–4024.
- Buzsáki G. 2006. *Rhythms of the brain*. Oxford University Press, New York.
- Cheng X, Crapse T, Tsao DY. 2013. Features that drive face cells: A comparison across face patches. In *Society for Neuroscience Conference*, San Diego, CA.
- Dubois J, de Berker AO, Tsao DY. 2015. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci* **35**: 2791–2802.
- Freiwald WA, Tsao DY. 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330**: 845–851.
- Grimaldi P, Saleem KS, Tsao DY. 2012. Anatomical connections of functionally defined anterior face patches in the macaque monkey. In *Society for Neuroscience Conference*, New Orleans, LA.
- Grimaldi P, Saleem KS, Tsao DY. 2013. Subcortical connections of the functionally-defined face patches in the macaque monkey. In *Society for Neuroscience Conference*, San Diego, CA.
- Issa EB, DiCarlo JJ. 2012. Precedence of the eye region in neural processing of faces. *J Neurosci* **32**: 16666–16682.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* **17**: 4302–4311.
- Kornblith S, Cheng X, Ohayon S, Tsao DY. 2013. A network for scene processing in the macaque temporal lobe. *Neuron* **79**: 766–781.
- Lafer-Sousa R, Conway BR. 2013. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat Neurosci* **16**: 1870–1878.
- Logothetis NK. 2008. What we can do and what we cannot do with fMRI. *Nature* **453**: 869–878.
- Moeller S, Tsao DY. 2011. Representation of face familiarity in AM. In *Society for Neuroscience Conference*, San Diego, CA.
- Moeller S, Freiwald WA, Tsao DY. 2008. Patches with links: A unified system for processing faces in the macaque temporal lobe. *Science* **320**: 1355–1359.
- Ohayon S, Freiwald WA, Tsao DY. 2012. What makes a cell face selective? The importance of contrast. *Neuron* **74**: 567–581.
- Ohki K, Chung S, Kara P, Hubener M, Bonhoeffer T, Reid RC. 2006. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature* **442**: 925–928.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R. 2012. Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *NeuroImage* **63**: 723–741.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R. 2014. Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *J Neurosci* **34**: 95–111.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. 2005. Invariant visual representation by single neurons in the human brain. *Nature* **435**: 1102–1107.
- Schummers J, Yu H, Sur M. 2008. Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex. *Science* **320**: 1638–1643.
- Sinha P. 2002. Qualitative representations for recognition. In *Lecture Notes in Computer Science*, pp. 249–262. Springer, New York.
- Tanaka K. 2003. Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cereb Cortex* **13**: 90–99.
- Tsao DY, Freiwald WA, Tootell RBH, Livingstone MS. 2006. A cortical region consisting entirely of face-selective cells. *Science* **311**: 670–674.
- Tsao DY, Moeller S, Freiwald WA. 2008a. Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci* **105**: 19514–19519.
- Tsao DY, Schweers N, Moeller SM, Freiwald WA. 2008b. Patches of face-selective cortex in the macaque frontal lobe. *Nat Neurosci* **11**: 877–879.
- Valentine T, Lewis MB, Hills PJ. 2015. Face-space: A unifying concept in face recognition research. *Q J Exp Psychol (Hove)* **2015**: 1–24.
- Viola P, Jones M. 2001. Rapid object detection using a boosted cascade of simple features. In *Computer vision and pattern recognition*. IEEE, Piscataway, NJ.
- Zhou H, Friedman HS, von der Heydt R. 2000. Coding of border ownership in monkey visual cortex. *J Neurosci* **20**: 6594–6611.



Cold Spring Harbor Symposia on Quantitative Biology

The Macaque Face Patch System: A Window into Object Representation

Doris Tsao

Cold Spring Harb Symp Quant Biol 2014 79: 109-114 originally published online May 5, 2015

Access the most recent version at doi:[10.1101/sqb.2014.79.024950](https://doi.org/10.1101/sqb.2014.79.024950)

References

This article cites 22 articles, 13 of which can be accessed free at:
<http://symposium.cshlp.org/content/79/109.full.html#ref-list-1>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

To subscribe to *Cold Spring Harbor Symposia on Quantitative Biology* go to:
<http://symposium.cshlp.org/subscriptions>
