# Journal Pre-proof


Does finding a face cell tell us anything much at all?

Doris Y. Tsao

PII: S0301-0082(20)30180-5

DOI: https://doi.org/10.1016/j.pneurobio.2020.101925

Reference: PRONEU 101925

To appear in: *Progress in Neurobiology*

Please cite this article as: { doi: https://doi.org/

**Does finding a face cell tell us anything much at all?**

**Doris Y. Tsao**[1,2*]

1. Division of Biology and Biological Engineering, Computation and Neural Systems, Caltech, Pasadena CA 91125.
2. Howard Hughes Medical Institute, Pasadena, CA 91125.

***Corresponding author:**
Doris Y. Tsao
Professor of Biology Investigator, HHMI
California Institute of Technology
Division of Biology
MC 114-96 Pasadena, CA 91125
USA

Phone: 626-395-1702 Fax: 626-395-8826
Email: dortsao@caltech.edu

There are two approaches to doing science. One is the "tractor" approach. You take a big, powerful piece of machinery and apply it in systematic fashion to a problem. Here, "you" is often a large group of people who share the same goal. A recent example is the International Brain Lab (Abbott et al., 2017), a consortium of labs across the world all performing the same experiment to understand visually-guided decision making in the rodent. The plan is for each lab to repeat the behavioral experiment while inserting advanced high channel count electrodes into different parts of the brain, like a fleet of tractors mowing a field.

A very different, older approach is that of the lone hunter pursuing a question no one else cares about, guided by a vision in his or her own head. Auden captures the essence of this approach in his wonderful poem "History of Science." The poem tells the tale of the Fourth Brother, who has been excised from the official fairy tale:

*Few of a sequel, though, have heard:*
*Uneasy pedagogues have censored*
*All written reference to a brother*
*Younger than the Third.*

*Soft-spoken as New Moon this Fourth,*
*A Sun of gifts to all he met with,*
*But when advised 'Go South a while!',*
*Smiled 'Thank You!' and turned North,*

*Trusting some map in his own head,*
*So never reached the goal intended*
*(His map, of course, was out) but blundered*
*On a wonderful instead,*

*A tower not circular but square,*
*A treasure not of gold but silver:*
*He kissed a shorter Sleeper's hand*
*And stroked her raven hair*

Charlie Gross was the Fourth Brother. He didn't care what others thought. At a time when the neuroscience community was celebrating the remarkable discoveries of Hubel and Wiesel on the mechanisms for early visual processing, Gross turned away from the crystalline chambers of V1, to a much murkier territory, inferotemporal (IT) cortex. In his History of Neuroscience autobiography, Gross explains why he started recording in IT cortex (Squire, 2009):

> "*Discouraged by my inability to understand the frontal lobe, I decided it lay in an inaccessible limbo bearing little relationship to anatomy, physiology, and psychology. …So I decided to turn my attention to the cortex on the inferior convexity of the temporal lobe.*"

I love this explanation: it is so very honest. We all have enchanted visions of what we will accomplish, but when we actually get into the lab, things become difficult, cells don't respond as we hoped they would. We face confusion and darkness, and sometimes, in the interest of making progress, we may make a decision as radical as to record in a new brain area simply because we have a hunch it will be more understandable--while still letting us address the original mysteries that captivated us (in Gross's case, higher-order vision and cognition).

In inferotemporal cortex, Gross found something much stranger than a tower square instead of circular: single cells that responded selectively to specific complex forms like a hand, a tree, or a face. He recounts the discovery of the first hand cell in his 1972 paper, "Visual Properties of Neurons in Inferotemporal Cortex of the Macaque" (Gross et al., 1972):

> *"…One day when, having failed to drive a unit with any light stimulus, we waved a hand at the stimulus screen and elicited a very vigorous response from the previously unresponsive neuron. We then spent the next 12 hr testing various paper cutouts in an attempt to find the trigger feature for this unit. When the entire set of stimuli used were ranked according to the strength of the response that they produced, we could not find a simple physical dimension that correlated with this rank order. However, the rank order of adequate stimuli did correlate with similarity (for us) to the shadow of a monkey hand."*

Reading this passage, one is inevitably reminded of another marathon electrophysiology session, when the edge of a slide was presented to a recalcitrant cell and found to elicit vigorous spikes. However, no Nobel Prize awaited Charlie Gross. The difficulty of systematically rank ordering effective stimuli for IT cells would require the advent of modern computers and the deep network revolution to begin to overcome (Yamins et al., 2014). Instead, what awaited Charlie Gross was mockery. Some people did not believe his findings of hand cells and face cells. And some, such as David Marr, argued that even if they were true, they were useless and un-illuminating. David Marr wrote in "Vision" (Marr, 1982):

> *"Suppose for example, that one actually found the apocryphal grandmother cell. Would that really tell us anything much at all? It would tell us that it existed—Gross's hand-detectors tell us almost that—but not <u>why</u> or even <u>how</u> such a thing may be constructed from the outputs of previously discovered cells…finding a hand-detector certainly did not allow us to program one."*

What Marr did not realize was that the face cells Gross discovered are actually clustered in IT into six anatomically connected patches (Tsao et al., 2003). By recording from each of these patches, it is in fact possible to gain significant insight into both why and how a face cell is constructed, as I explain below. Moreover, Marr did not foresee that understanding the gross structure of the macaque ventral visual pathway, a hierarchical network beginning with center-surround ganglion cells in the retina and culminating in neurons with "fully-connected" receptive fields selective for specific complex forms in IT cortex, would in fact enable a hand-detector to be programmed successfully for the first time (Krizhevsky et al., 2017).

Using functional magnetic resonance imaging (fMRI) in monkeys, we found six patches of cortex that responded much more strongly to faces compared to other objects in IT (Tsao et al., 2003, 2008). These six patches are distributed along the entire length of IT cortex. Interestingly, one of them, MF, corresponds precisely to the anatomical site (fundus of the STS, 5-7 mm anterior to the ear canals) where Gross and his colleagues had found the highest percentage of face-selective cells reported prior to our work (34% of recorded cells) (Desimone et al., 1984). Cells in the middle face patches ML and MF respond selectively to specific face views, cells in face patch AL respond invariantly to mirror symmetric profile views or to downwards/upwards/frontal views, and cells in the most anterior patch AM respond to faces invariantly across all views (Freiwald and Tsao, 2010)f. This functional progression immediately suggests a wiring scheme in which face cells at each stage pool specific sets of inputs from the preceding stage. Various computational schemes have been advanced to explain the precise rules governing the hierarchical pooling (Leibo et al., 2017; Yildirim et al., 2020), and the jury is still out. Nevertheless, we can confidently say that single-unit recordings of face cells have shed substantial light on how a face cell is wired.

These single-unit recording studies have also shed light on why face cells exist. My lab recently found that face cells in ML/MF and AM are well-modeled as computing a linear projection of incoming faces formatted in "shape-appearance" coordinates onto a specific preferred axis (Chang and Tsao, 2017). Because the shape-appearance model is a graphical model for generating faces, this means that the code harbored by face patches enables not only face discrimination but also *face generation*. The idea that cells at the highest stage of IT cortex carry enough information to re-generate the input is a deep insight. It suggests that the purpose of IT cortex is to generate a high-fidelity representation of objects that can be used for any face-related task, not just face identification—addressing why face cells exist.

Gross was truly a scientist ahead of his time. His remarkable discovery that cells in IT cortex are selective for specific visual forms such as faces, hands, and trees is only now beginning to be truly appreciated by the scientific community. I remember running into David Hubel's office during graduate school to tell him about a face cell I was recording from. David came and listened to the cell together with me, and we both marveled at how it responded only to faces. Afterwards, he told me how sorry he felt for how he had treated Charlie, dismissing his findings--now he could appreciate their importance. Since I was just a naïve graduate student, this apology flew over my head. But now I realize, the person he was really speaking and apologizing to was Charlie.

Of course, it makes sense why scientists were initially so skeptical of Gross's findings. It *does* seem incredible that the brain should contain cells specialized for faces. Does that imply the brain should have a cell for every object one encounters (on my desk, I see: a barrette, a tape measure, a candle, a cup of tea, a violin shoulder pad, a coaster….)? Wouldn't one run out of cells? Gross explains the answer to this conundrum with crystal clarity:

> *"Are the face and hand cells found in IT cortex examples of the "grandmother cells" of Lettvin (in Barlow, 1995), cells that respond only to a specific visual concept, such as your own grandmother "however displayed, whether animate of stuffed, seen from behind, upside down, or on a diagonal, or offered by caricature, photograph or abstraction"? Are they examples of the "gnostic" cells of Konorski (1967), neurons that represent "unitary perceptions"? The available evidence provides an overwhelming "no" to both possibilities. IT cells that respond only to a specific object, such as the face of one individual, and continue to do so across various transformations have never been seen. Rather IT face cells respond in varying degrees to a set of faces and never solely to one. Different IT cells show a different pattern of responses to a set of faces. Thus the coding of faces (and presumably other objects) appears to be done by the pattern of firing over a set of cells, that is, by what has been termed "coarse coding," "ensemble coding," "population coding," or "cross-fiber pattern coding." This absence of one cell–one visual concept is true for both natural stimuli such as faces as well as for arbitrary stimuli that evoke responses of IT cells after explicit training (Gross, 1992, 2002)."*

In some way, almost all of my work to date has been simply clarifying this one paragraph from Charlie Gross. As I already mentioned, we now understand fairly well the IT face code for facial identity, and it turns out to be extremely simple, with single cells acting as rulers projecting incoming faces onto specific axes of face space (Chang and Tsao, 2017). This simple model can account for 80% of the explainable variance of face cell responses to a set of randomly generated synthetic human faces, and allows reconstruction of such faces with remarkable accuracy. Interestingly, the computation of axis projection predicts that each face cell has an infinite number of metamers, distinct faces that elicit an identical response in the cell, since all the faces in the hyperplane orthogonal to the preferred axis should project to the same value. We confirmed this experimentally. Hence Gross was absolutely right, IT face cells are not grandmother cells.

For me, the most remarkable proof of the utility of face cells for understanding the brain comes from a recent study from my lab seeking to understand the overall organization of IT cortex using the same tools that we used to understand the face patch system. By electrically microstimulating various parts of IT cortex, performing fMRI, and targeting electrodes to specific fMRI-identified areas, we discovered that roughly half of IT cortex can be described as a coarse topographic map of a 2D object space, repeated three times (Bao et al., 2020). The axes of this object space can be computed using a deep network. Two quadrants of this object space turn out to coincide with faces and bodies, while the other two do not correspond to any easily nameable visual category; instead, objects in these remaining two quadrants share a visual shape property (spiky in one quadrant, stubby in another). What this result tells us is that the face patch network is not a unique structure within IT cortex, but instead possesses multiple siblings. Furthermore, each sibling network shares the same anatomical organization and coding principles as the face patch network. In particular, there is increasing view invariance as one proceeds anterior within each network, and cells within each network are projecting objects onto specific axes of object space. This work reveals the face patch system to be a neural Rosetta Stone, enabling us to translate our understanding of faces to understanding of a much larger swath of IT cortex.

Looking to the future, a key question is how IT information is read out. What happens next? How are IT object representations incorporated into a sensible model of the world that allows us to act and to build meaningful memories? Here, we are still in darkness, and almost certainly fundamentally new concepts will be necessary. Hence, as we read articles pronouncing that *“[Science] is no longer driven by lone figures labouring in their laboratories, but has become a team effort that spans labs, departments, disciplines, institutions and continents. …[I]t often relies now on data sets so vast that human brains cannot hope to hold or parse them all” (Ball, 2019),* let us not forget the deep lessons of Charlie’s scientific journey: Don’t be afraid to go into the desert alone (claustrum? pulvinar? frontal pole?). And don’t ignore the next face cell--however strange it may seem and however much your colleagues may ridicule you for it.

I close with a fond memory I have of Charlie. A few years ago, John Allman and I invited him and Joyce to come to Caltech to participate in a conversation about art and science. The day after the event (which was a huge success, drawing a capacity crowd to Baxter Auditorium), Charlie, Joyce, John, and I went on a hike in the San Gabriels. I remember Charlie finding a wooden stick on the ground and using it as a cane. And I remember how warm and generous he was, asking about my life and my lab and sharing with me stories about his experiences teaching neuroscience to prison inmates at San Quentin. As he rapidly descended down a steep pass with his wooden cane, the vulnerability and love between Joyce and Charlie became exposed when she expressed her concern for his safety, "What will I do if you fall, Charlie?" I felt deeply grateful for this opportunity to have a personal interaction with someone who was a larger-than-life figure in my imagination, and who had been my longtime guide and hiking partner in IT cortex.

**Acknowledgments:** I thank Janis Hesse for comments on the manuscript.

Charlie riding a camel in the desert.

**References**

Abbott, L.F., Angelaki, D.E., Carandini, M., Churchland, A.K., Dan, Y., Dayan, P., Deneve, S., Fiete, I., Ganguli, S., Harris, K.D., Häusser, M., Hofer, S., Latham, P.E., Mainen, Z.F., Mrsic-Flogel, T., Paninski, L., Pillow, J.W., Pouget, A., Svoboda, K., Witten, I.B., Zador, A.M., 2017. An International Laboratory for Systems and Computational Neuroscience. Neuron 96, 1213–1218. https://doi.org/10.1016/j.neuron.2017.12.013

Ball, P., 2019. Science must move with the times. Nature 575, 29–31. https://doi.org/10.1038/d41586-019-03307-8

Bao, P., She, L., McGill, M., Tsao, D.Y., 2020. A map of object space in primate inferotemporal cortex. Nature 583, 103–108. https://doi.org/10.1038/s41586-020-2350-5

Chang, L., Tsao, D.Y., 2017. The Code for Facial Identity in the Primate Brain. Cell 169, 1013-1028.e14. https://doi.org/10.1016/j.cell.2017.05.011

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

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

Gross, C.G., Rocha-Miranda, C.E., Bender, D.B., 1972. Visual properties of neurons in inferotemporal cortex of the Macaque. J. Neurophysiol. 35, 96–111. https://doi.org/10.1152/jn.1972.35.1.96

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386

Leibo, J.Z., Liao, Q., Anselmi, F., Freiwald, W.A., Poggio, T., 2017. View-Tolerant Face Recognition and Hebbian Learning Imply Mirror-Symmetric Neural Tuning to Head Orientation. Curr. Biol. 27, 62–67. https://doi.org/10.1016/j.cub.2016.10.015

Marr, D., 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, Vision. The MIT Press.

Squire, L.R., 2009. The History of Neuroscience in Autobiography Volume 6, The History of Neuroscience in Autobiography Volume 6. Oxford University Press.

Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., Tootell, R.B.H., 2003. Faces and objects in macaque cerebral cortex. Nat. Neurosci. 6, 989–995. https://doi.org/10.1038/nn1111

Tsao, D.Y., Moeller, S., Freiwald, W.A., 2008. Comparing face patch systems in macaques and humans. Proc. Natl. Acad. Sci. U.S.A. 105, 19514–19519. https://doi.org/10.1073/pnas.0809662105

Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. U.S.A. 111, 8619–8624. https://doi.org/10.1073/pnas.1403112111

Yildirim, I., Belledonne, M., Freiwald, W., Tenenbaum, J., 2020. Efficient inverse graphics in biological face processing. Sci Adv 6, eaax5979. https://doi.org/10.1126/sciadv.aax5979