Cognitive Neuroscience
9/7/21

Pattern Vision

1. Human neuropsychology and imaging
2. Image selective neurons in monkey inferotemporal cortex
3. Limited invariance across changes of viewpoint
4. Inter-image distance in multi-neuronal activation space
5. Inferotemporal activity in relation to perception

IT → ventral stream terminus

TE → occupies bulk of IT cortex

KEY: I still don't know.

PIG: Could be a dog or any other animal.

BIRD: Could be a beach stump.

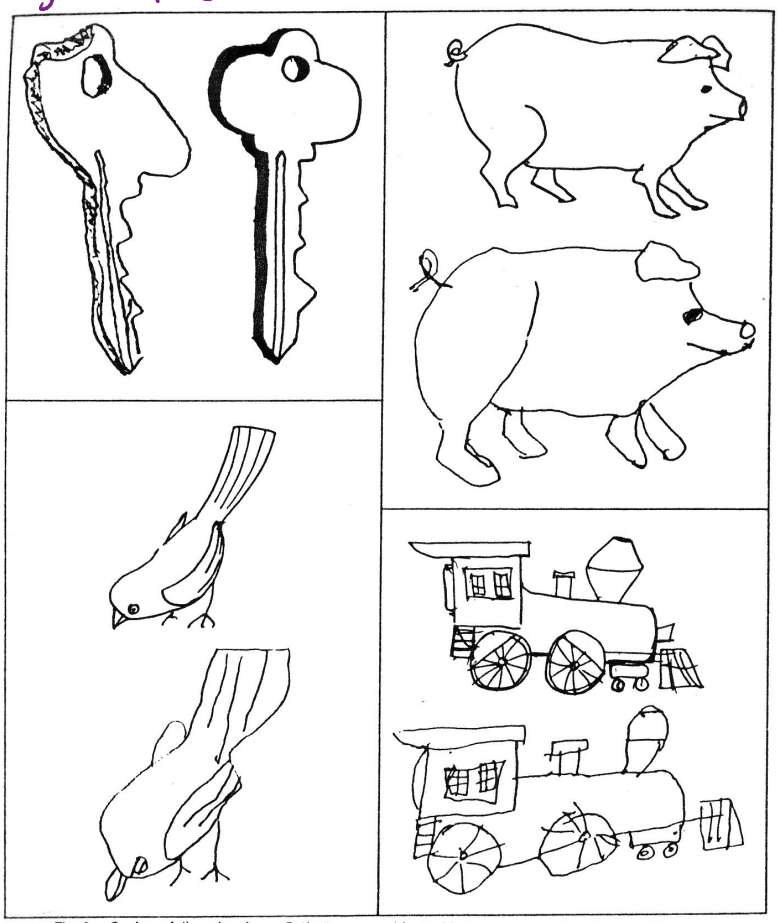
LOCOMOTIVE: A wagon or a car of some kind. The larger vehicle is being pulled by the smaller one.

good copies

Clearly not blind

Doesn't know what he drew

Can identify that it is a four-legged animal



Beach stump?

Thinks it is a vehicle

Fig 4.—Copies of line drawings. Patient was unable to identify any before copying. After making copy, his identifications were **top left,** key—"I still don't know"; **top right,** pig—"Could be a dog or any other animal"; **bottom left,** bird—"Could be a beach stump"; **bottom right,** locomotive—"A wagon or a car of some kind. The larger vehicle is being pulled by the smaller one."

Not impaired in recognizing other sensory modalities

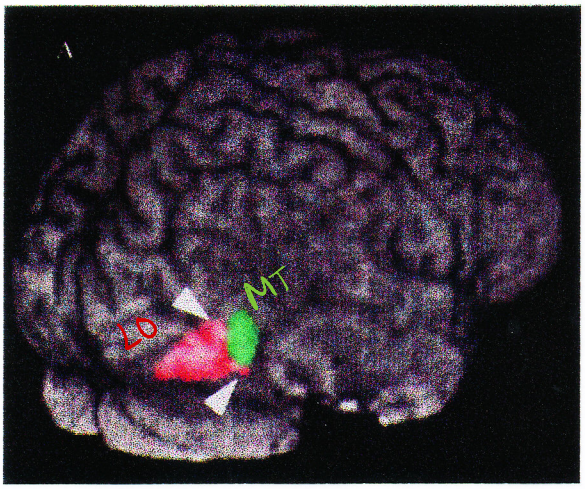
Patients with visual object agnosia can see and even copy accurately, and yet are poor at recognizing images - an indication that recognition is dissociable from other aspects of vision.

Rubens & Benson
Arch. Neurol. 24:
305-316 (1971).

Impairment in object recognition

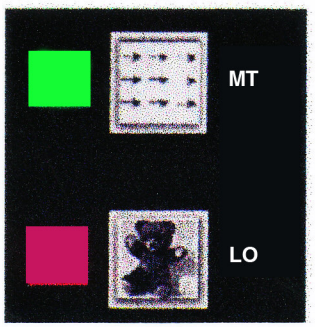
Neurobiology: Malach *et al.*
Proc. Natl. Acad. Sci. USA 92 (1995)

Area LO



objects vs. textures

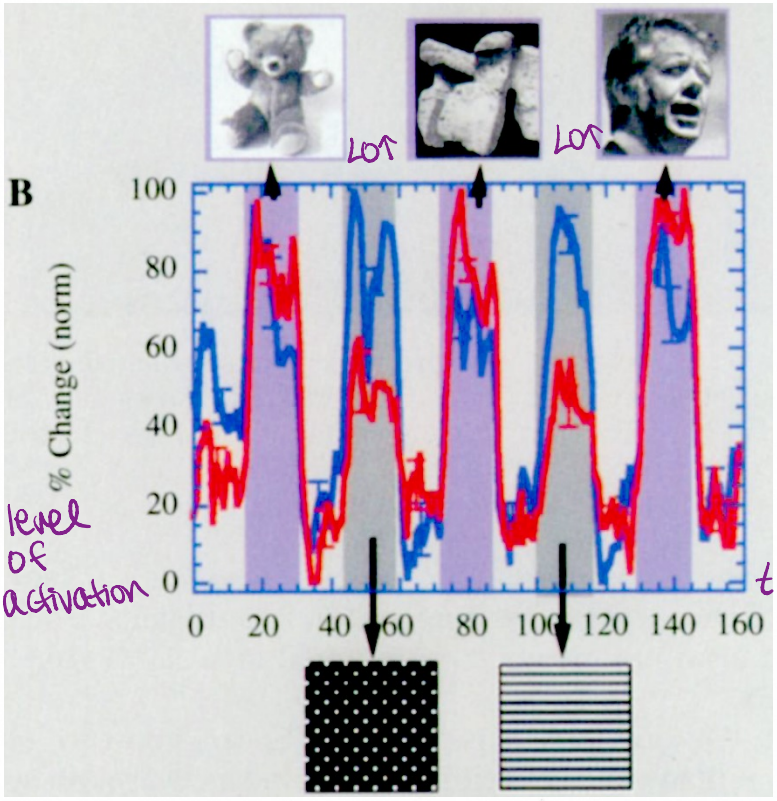
Malach et al., in a 1995 functional imaging study in human subjects, discovered that a "lateral occipital" area (LO: red in top figure) located posterior to area MT (green in top figure), responded more strongly to pictures of objects than to textures. An example of the supporting data is shown in the bottom figure. The red trace represents blood flow in area LO whereas the blue trace represents blood flow in primary visual cortex (area V1). Note that when subjects were viewing photographs of objects, area LO was more active than area V1 whereas V1 was more active during viewing of textures.

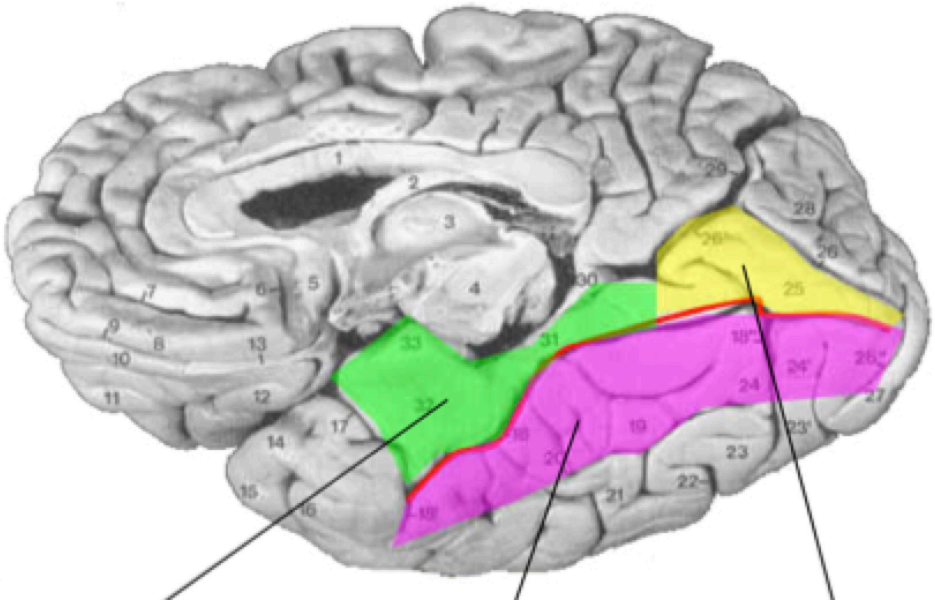


Anatomically juxtaposed in human cortex

Primary visual
Area V1

lateral occipital
Area LO





Parahippocampal Gyrus | Fusiform Gyrus | Lingual Gyrus

Condition	Impaired	Main Locus	fMRI Correlate
Object agnosia	Objects	Bilateral Temporal	Temporal Object > Texture
Prosopagnosia	Faces	Right Medial Temporal	Fusiform Face Area
Pure alexia	Words (letter-by-letter reading)	Left Medial Temporal	Visual Word Form Area
Topographic amnesia	Places (tend to get lost)	Medial Temporal	Parahippocampal Place Area

The unifying theme of the impairments arising from damage to the indicated distrrict of temporal lobe cortex is that they involve visual recognition.

subtle differentiation between details in a category (maybe not just faces?)

Prosopagnosia

Patients typically cannot:

Recognize acquaintances or family members

Recognize themselves in mirrors

Recognize famous individuals in photographs

Some patients have been reported to be unable to:

Recognize individual cows within a herd

Recognize species of tropical fish

Recognize jars of foodstuff in a grocery store

Patients may fail when asked to make fine discriminations:

Report whether two unfamiliar faces are the same

Distinguish age, sex and emotion of a face

Patients typically can make coarse discriminations:

Recognize a face as a face

Name and point to parts of a face

Patients compensate by:

Recognizing people from their voices

Recognizing people from distinctive hairstyles, glasses, etc.

Prosopagnosia may be congenital:

It tends to run in families

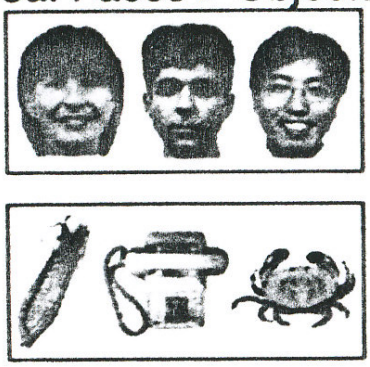
Here is a first-person account by a Washington Post columnist:

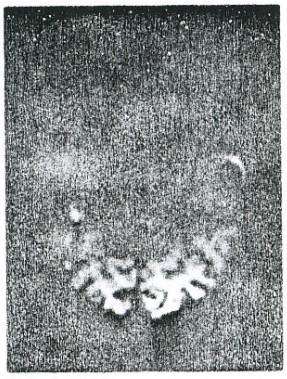
Last year, I was trailing behind Steve — now my husband — in a grocery store when he grabbed a jar of store-brand peanut butter from a shelf. I plucked it out of our cart and examined the label. "Since when do you buy generic?" I demanded. Steve jumped away from me, his eyes wide with fear and surprise. It was an expression unlike anything I'd seen cross my husband's face before — because, I belatedly realized, this man was not my husband. I dropped the peanut butter jar and sprinted off — leaving this poor stranger utterly perplexed. When I found Steve in the frozen-food aisle, I told him what had happened. "It's because you have the same coat," I explained.

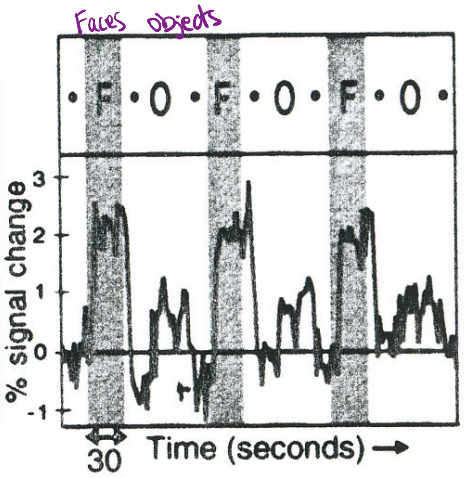
Sadie Dingfelder, My Life with Face Blindness, Washington Post, August 21, 2019

Fusiform Face Area

3a. Faces > Objects

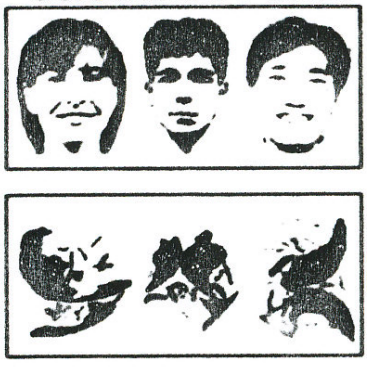


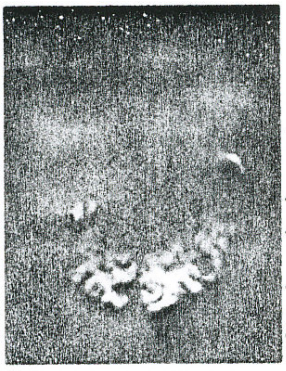


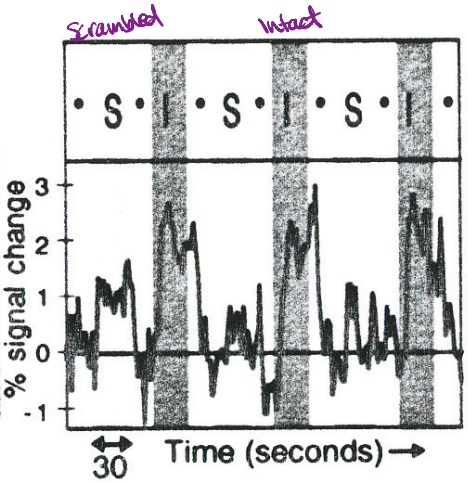


Faces > Houses

3b. Intact Faces > Scrambled Faces

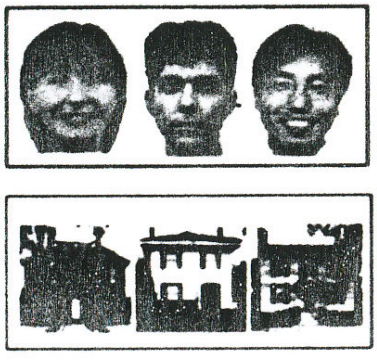


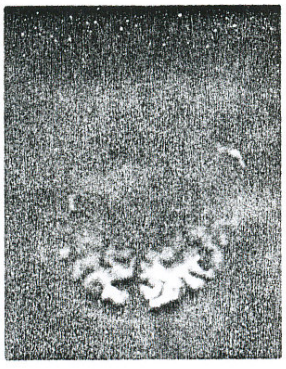


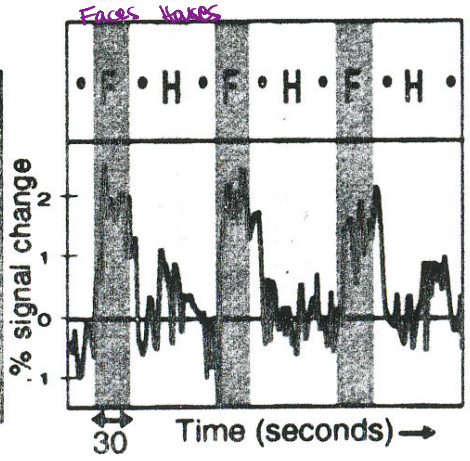


Intact > Scrambled

3c. Faces > Houses



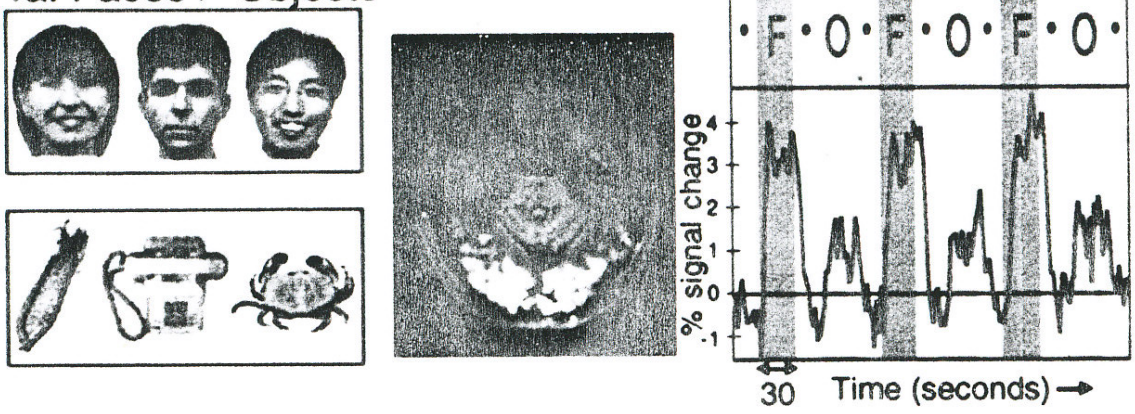




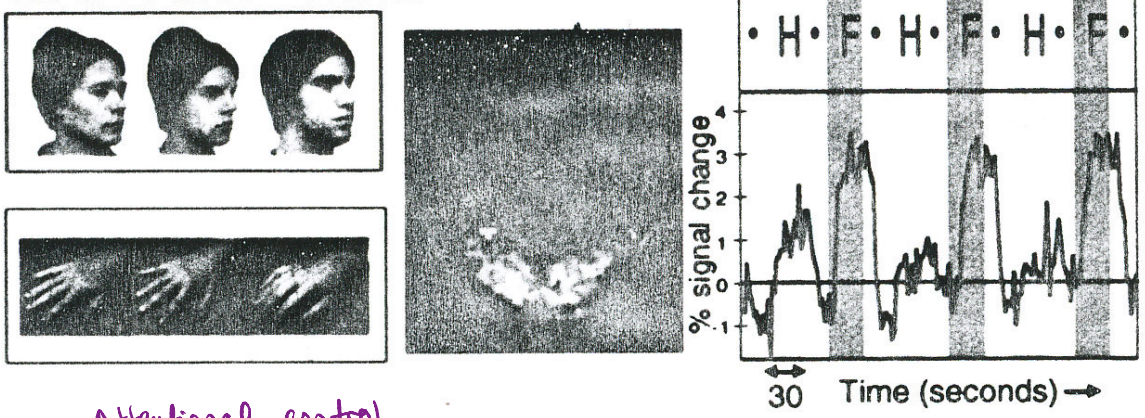
Faces > Houses

Fusiform Face Area

4a. Faces > Objects



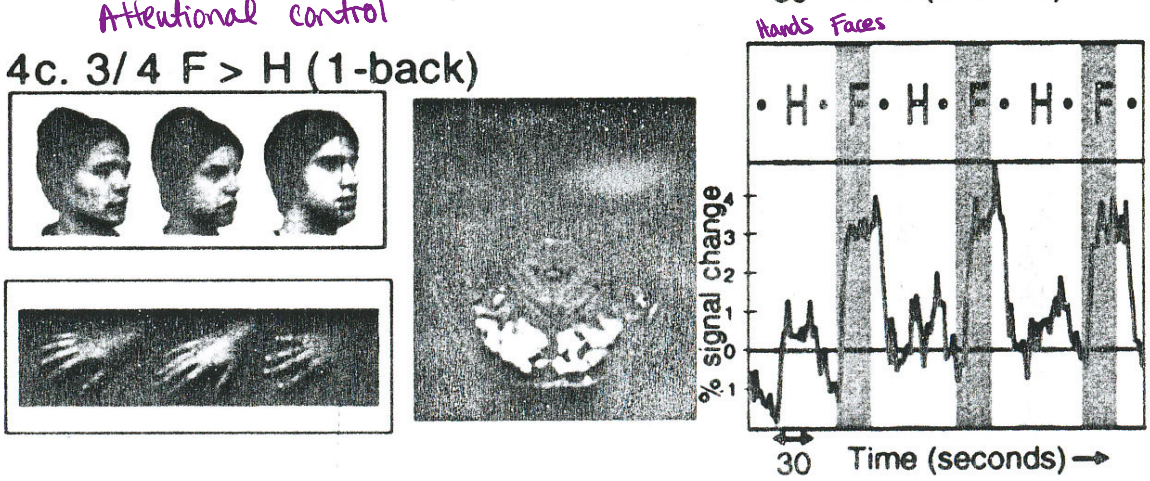
4b. 3/4 Faces > Hands



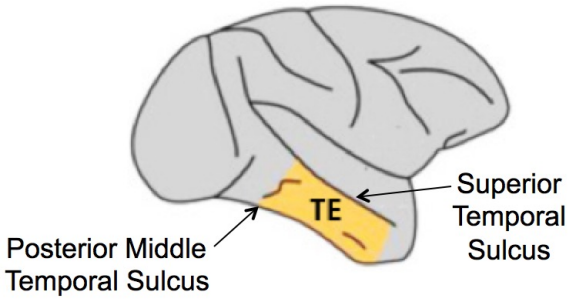
Faces > Hands

Attentional control

4c. 3/4 F > H (1-back)

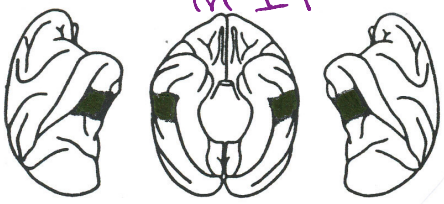


Faces > Hands



Temporal lobe lesions create impairment

Lesions bilaterally in IT



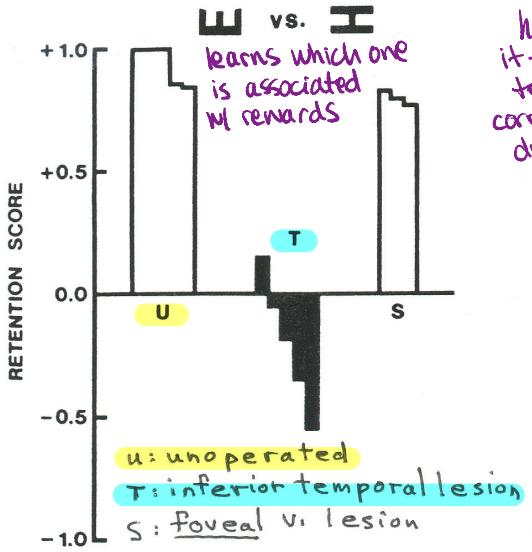


Figure 2. Retention scores (see the text) for the discrimination learned prior to operation for the unoperated (*U*), inferior temporal (*T*), and striate (*S*) groups. Each *bar* within a group represents a single animal. Within the IT group, from *left* to *right*, the individual *bars* represent animals IT-1 to IT-5. The discriminanda are shown at the *top* in reverse contrast and were 3.2 cm long.

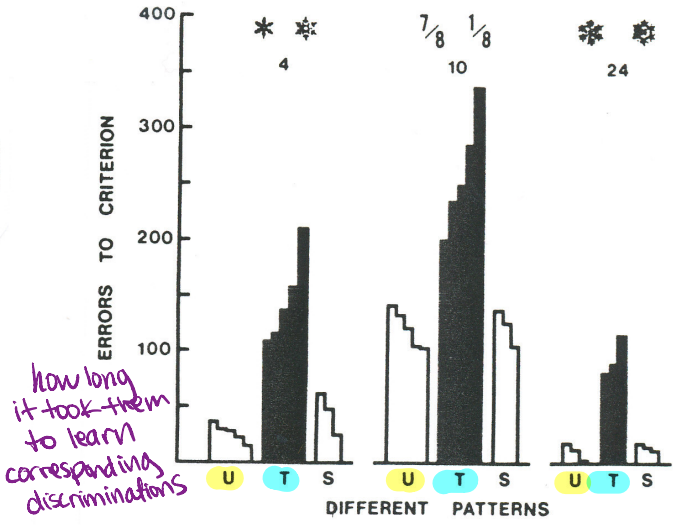


Figure 3. Errors made prior to reaching criterion on the problems on which the discriminanda were different patterns. The *problem number* and the *discriminanda* are shown *above* each set of *histograms*. The discriminanda are shown in reverse contrast on the same scale as those in Figure 2 (see also legend to Fig. 2). In this and subsequent figures the *bars* for the individual animals are shown in the same order as in Figure 2.

Demonstration of visual object agnosia in monkeys with lesions of inferotemporal cortex (T) as compared to control monkeys either unoperated (u) or with foveal scotomas induced by lateral striate cortex lesions (s).

Holmes & Gross
J. Neurosci. 4: 3063-3068
(1984)

Within each continuous group of bars, each bar represents one monkey trained on the illustrated discrimination

One day, 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 a previously unresponsive neuron. We then spent the next 12 h 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 did correlate with similarity (for us) to the shadow of a monkey hand.

Epiphany:

need to show complex objects to elicit response in IT cortex

Gross, Arch. Neuropsychologia 46: 841-852 (2008)

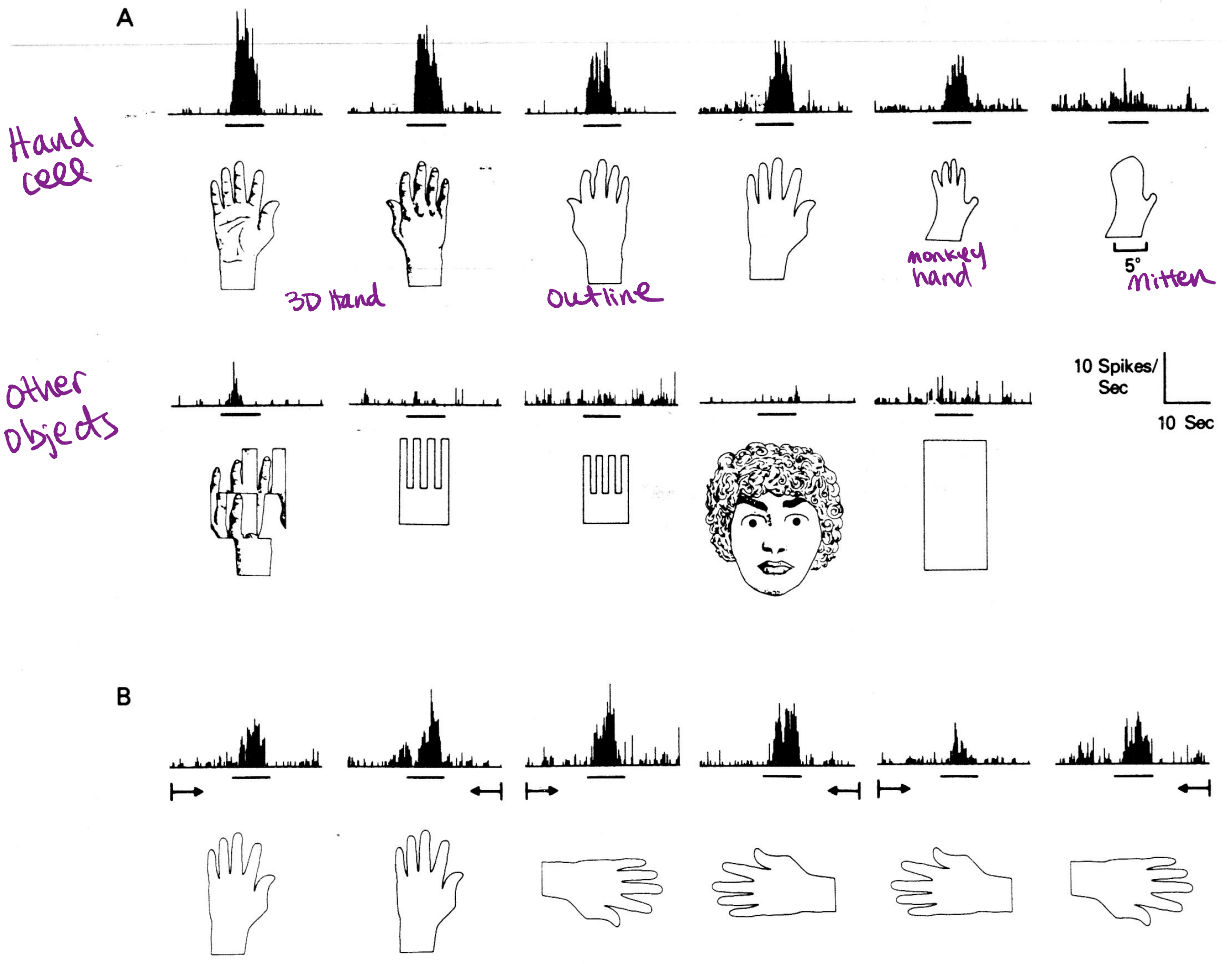


Figure 5. Responses of a unit that responded more strongly to hands than to any other stimulus tested. *Drawings* under each *histogram* were traced from the stimuli. *A*, Comparison of responses to hands versus other patterns. Stimuli were randomly interleaved and included the front and back of a model of a human hand, white cutouts with the same shape as the human hand model, a cutout with the shape of a monkey hand, a cutout of a monkey's hand with the space between the fingers eliminated, a scrambled photograph of the model of the human hand (10 rearranged pieces), two "grating-like" hands, a model of a human face, and a plain rectangle. Stimuli were moved at 1.2°/sec from the contralateral into the ipsilateral visual field and were visible within a 15° window centered on the fovea. *B*, Responses to a stimulus with the shape of a hand, in different orientations. The contralateral visual field is represented on the *left* of each *histogram* and the ipsilateral field on the *right*. The *arrows* indicate the direction of stimulus motion and the direction of time in the *histograms*. Other conditions were as in *A*.

Desimone et al., J. Neurosci. 4: 2051-2062 (1984)

Data from an early study in Charlie Gross' lab demonstrating neuronal selectivity for visual patterns in monkey IT cortex

Grandmother Cell Hypothesis

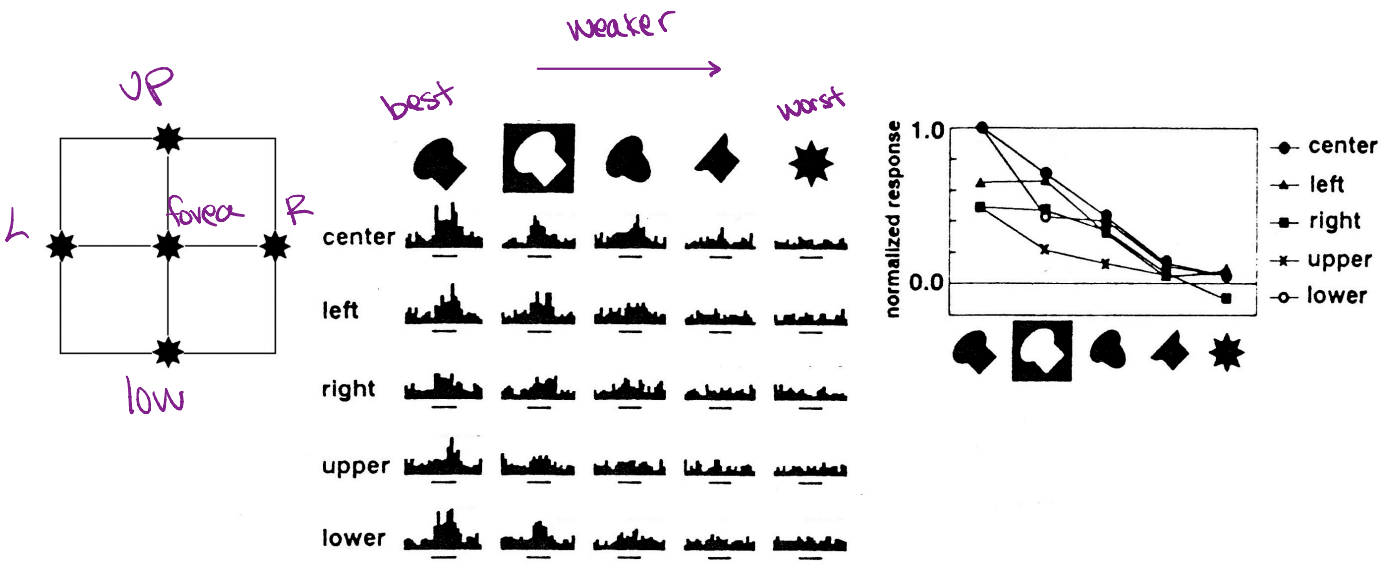
A cell responds only to a specific visual concept, such as your own grandmother however displayed, whether animate or stuffed, seen from behind, upside down, or on a diagonal, or offered by caricature, photograph or abstraction.

Responds to objects regardless of viewpoint

Viewpoint invariance

C.G. Gross, Neuropsychologia 46:841 (2008).

Location Invariance



Ito et al., J. Neurophysiol. 73: 218-226 (1995)

IT neuron show consistent object selectivity regardless of location

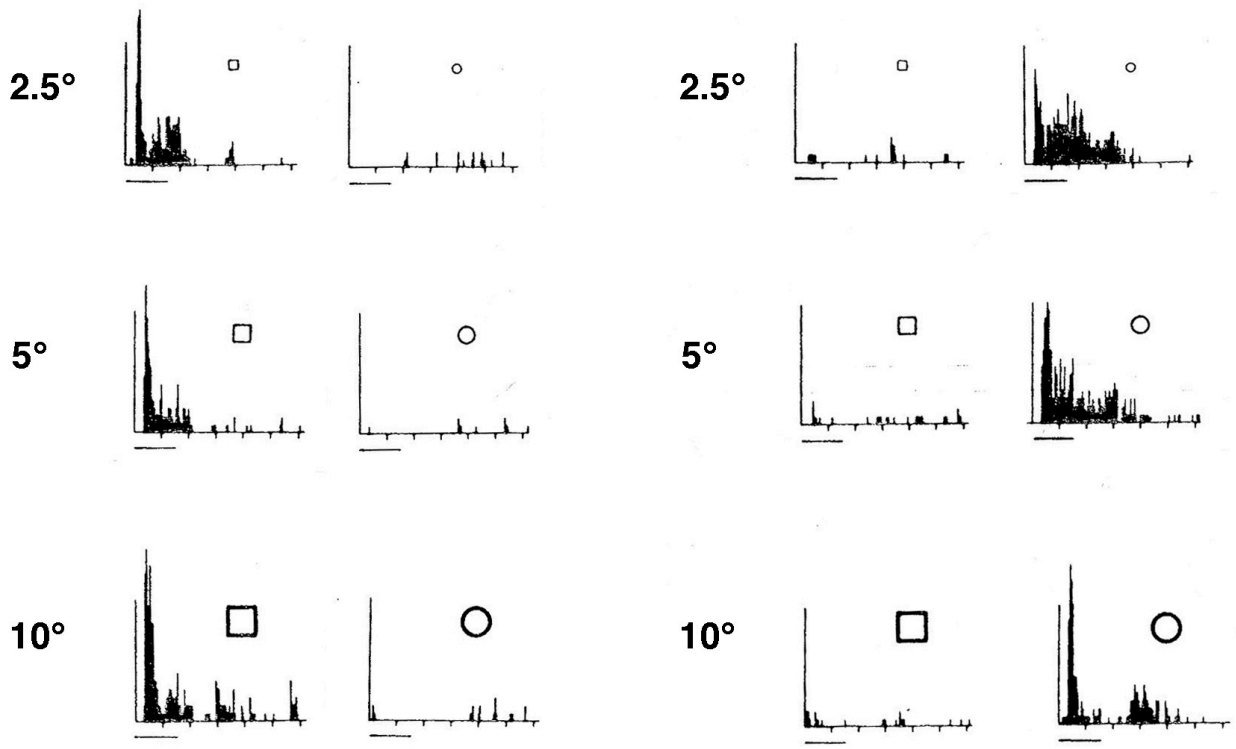
IT neuron show consistent object selectivity regardless of scale

(not perfect, pattern Δ's a bit)

Scale Invariance

Cell 1
size-invariant
square selectivity

Cell 2
size-invariant
circle selectivity

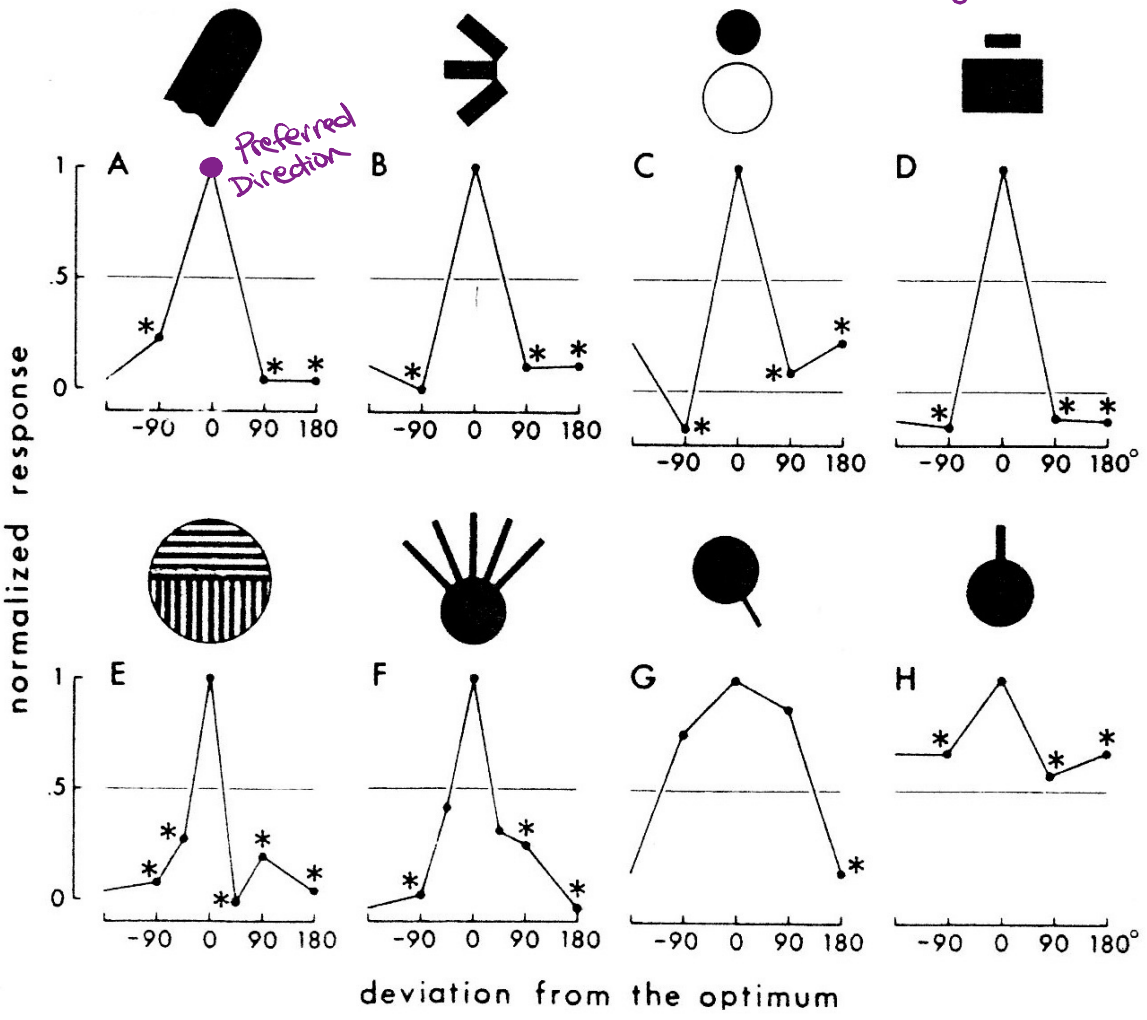


Sato et al., Exp. Brain Res. 38:313-319 (1980)

w/in screen orientation

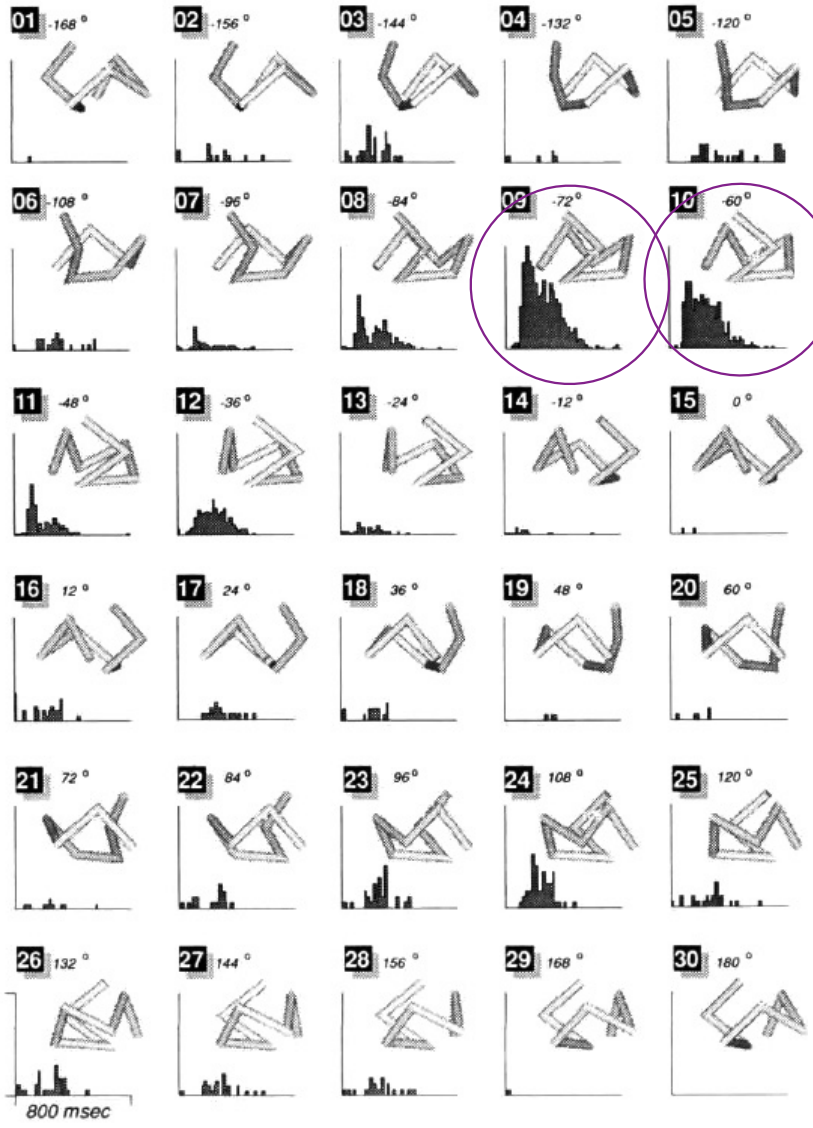
Little Orientation Invariance

As you rotate away from pref direction, FRs ↓



Tanaka et al., J. Neurophysiol. 66: 170-189 (1991)

Little Orientation Invariance

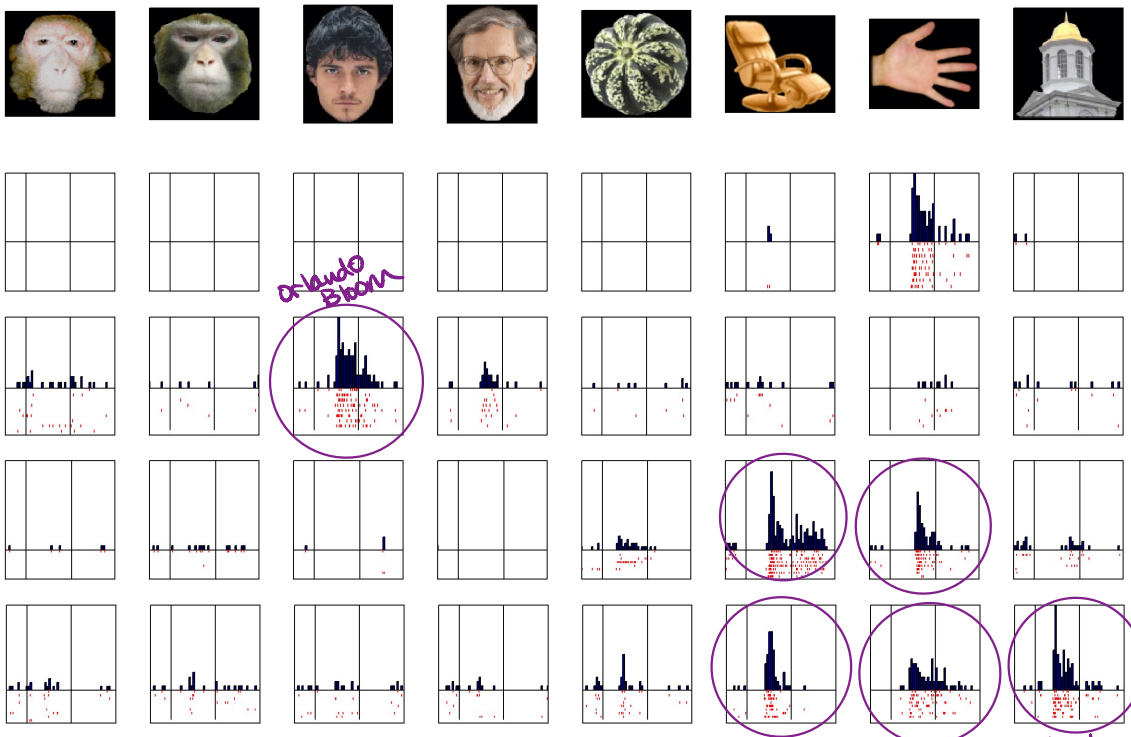


N.K.Logothetis and J. Pauls. Cereb. Cortex 5: 270 (1995).

IT don't represent grandmother, represent particular snapshot

From the absence of orientation invariance, we would be forced, following Lettvin's logic, to conclude that each IT neuron is selective for a particular snapshot of your grandmother. However there's a problem even with even this modest interpretation, namely that a neuron typically responds to many different images.

Atypical narrowly selective neurons.



Typical broadly selective neurons.

For each IT neuron, found object w/ strong response & then reduced object to bare bones

What is simplest stimulus that will make it fire?

Subtractive Analysis

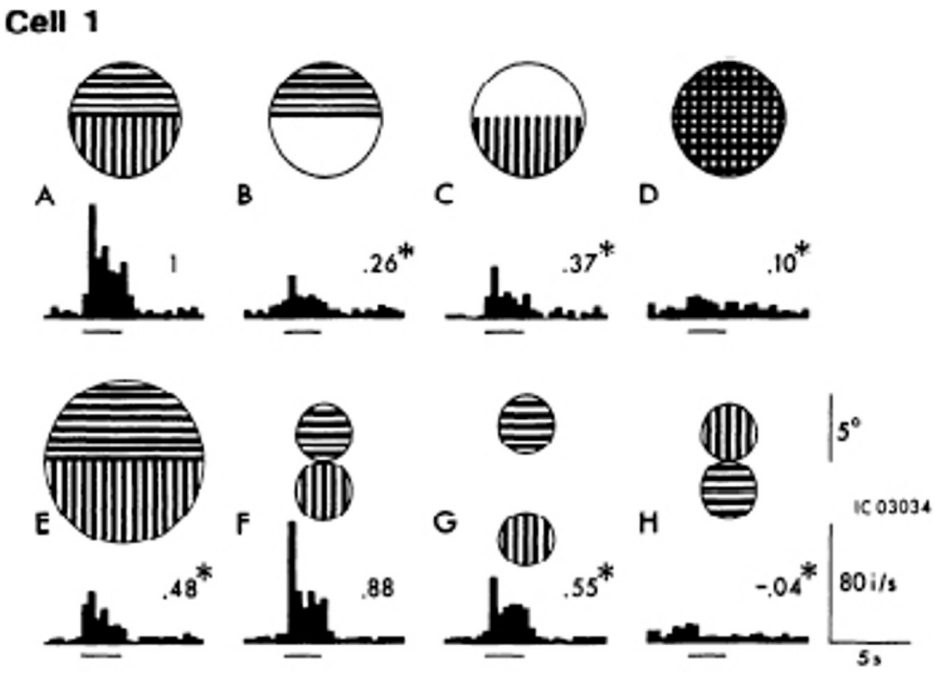
The routine set was composed of *1*) slits and spots of various sizes and colors, projected from a handheld projector; *2*) rectangular and circular paper cuts of various sizes and colors, presented in front of the screen; *3*) regular geometric patterns, such as stripes, dot patterns, concentric rings, and patterns like windmills, projected from a handheld projector; *4*) plastic and sponge spheres, sponge cubes, plastic cylinders, and feather brooms, of various colors; 5) 3-D animal imitations made of vinyl, cloth, or plastic, including imitations of tiger, tabby cat, spotted dog, zebra, giraffe, gorilla, hawk, duck, frog, raccoon, monkey, human head, and human hand; *6*) 3-D plant imitations made of plastic, including banana, apple, orange, maize, pineapple, grapefruit, melon, cabbage, carrot, potato, cucumber, watermelon, eggplant, onion, potato, a bunch of grapes, ivy, a potted plant, a cut piece of apple, and an obliquely cut sweet potato (which was fed every day to the monkeys in the cage); and *7*) the experimenter's hands, body, and face. Various sides of the objects were presented with various orientations.

If a cell gave consistent response to one of the 3-D objects, we tried to clarify which component or combination of components of the image were essential for the activation. If a cell was consistently activated by more than two different stimuli, we started from the features common to the stimuli. We made two-dimensional (2-D) paper models that simulated the image of the object and compared the response of the cell to these 2-D paper models and the original 3-D object. The paper model simulated not only the shape of the outline, but also the texture and the color of the image. If the cell responded to the 2-D model as strongly as to the 3-D object, we then reduced the complexity of the 2-D model step by step and assumed that the simplest 2-D feature that fully activated the cell was the feature that the cell extracted from the image.

Tanaka et al.
J. Neurophysiol. 66:170-189 (1991)

Subtractive Analysis

Finding simplest effective stimulus



Tanaka and colleagues tried heroically to identify the thing for which each IT neuron was selective by first finding a 3D object that elicited a strong response and then successively reducing it to simpler and simpler forms, using scissors, paper and felt tip pens.

Here are the results of subtractive analysis carried out on a neuron initially identified as responding strongly to a doll in the form of a tabby cat.

They referred to the simplest form that could drive a neuron strongly (Image A in this figure) as a partial feature, implying that it could occur as a component of multiple images and that the neuron was selective not for the tabby cat but for any image containing the partial feature.

Tanaka et al.
J. Neurophysiol. 66:170-189 (1991)

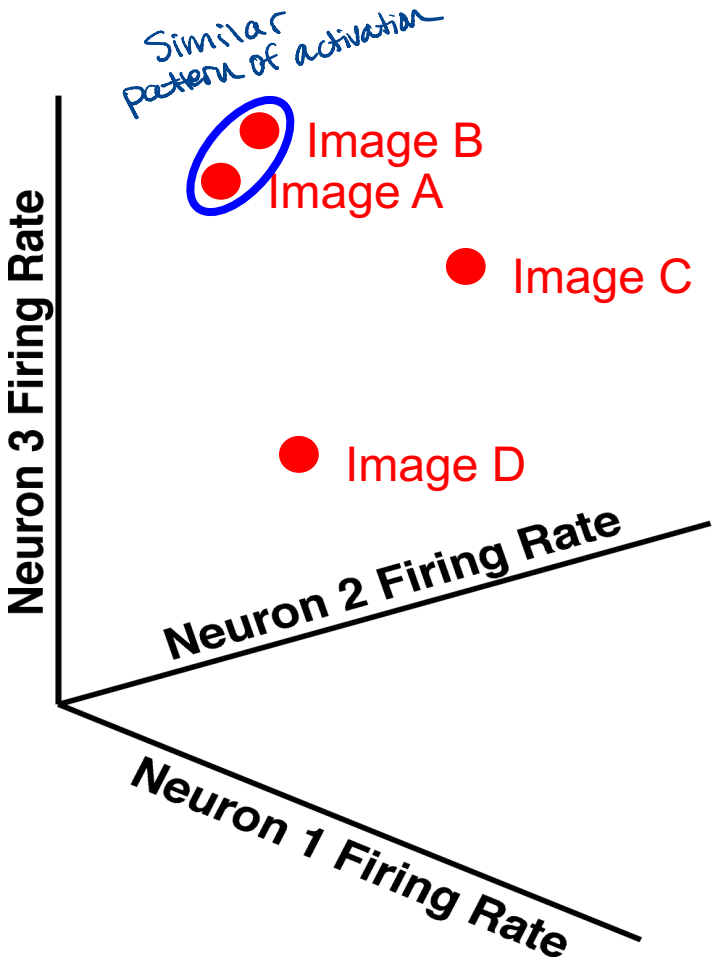
Combination coding

Although responses of anterior IT cells were thus selective to particular features, the features coded by individual cells were not complex enough to indicate a concept of particular real objects. Instead, they may be partial features common to images of several different objects. This means that simultaneous activation of a few to a few tens of cells is required to indicate the concept of a particular object. In addition, because the responses of the cells were selective for the orientation of the stimulus, different sets are required for indicating view of the object with different orientations. Images of objects are thus coded by combinations of active cells each of which represents the presence of a particular partial feature. We will call this type of coding "combination coding." An advantage of the combination coding over the "local coding" (Barlow 1972) is the capability of generalization, although not so strong as that of the "distributed coding." That is, knowledge acquired for an item represented by a population of cells is automatically generalized for other items represented by mostly overlapping populations (Hinton et al. 1986).

Can rate images for similarity

Commonality of neurons firing for objects

Tanaka et al.
J. Neurophysiol. 66:170-189 (1991)



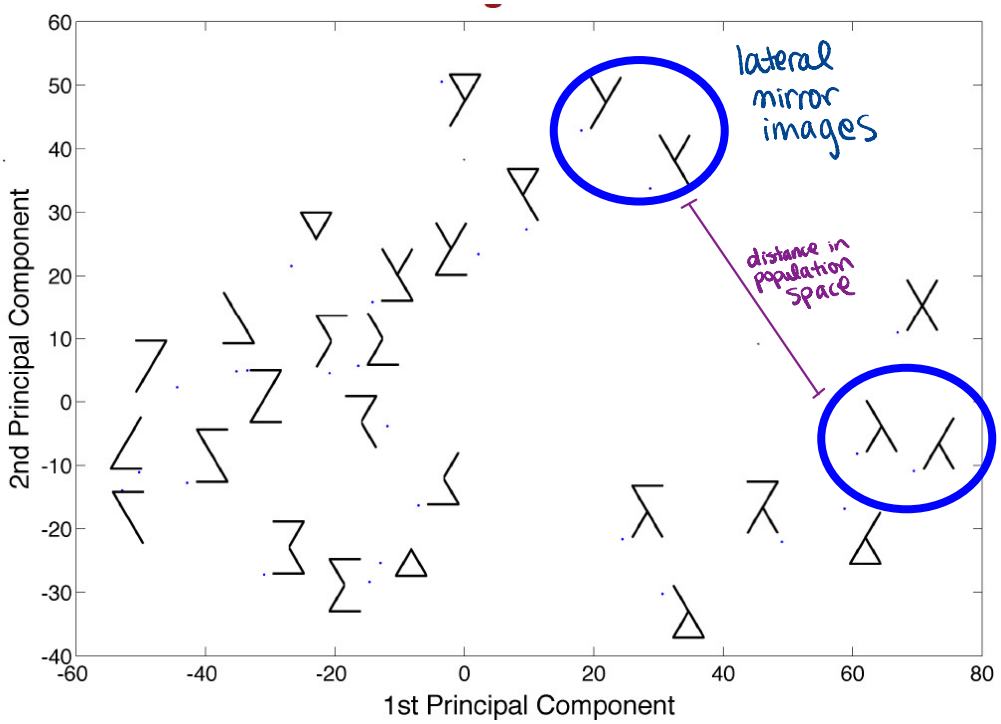
If each IT neuron represents a partial feature potentially present in many images (as suggested by Tanaka) and if each neuron fires in proportion to how closely the content of the current image matches its partial feature, then the representation of images in IT is distributed (each image is represented by the activity of a great many neurons) and continuous (it matters not which neurons are active but how much each neuron is active).

Think of each image as a point in a multi-neuronal activation space with the position of the point determined by how strongly the image makes each neuron fire. Points close together represent images eliciting similar patterns of population activity. This seems to be more about representing things for recognition than about recognition per se.

What does it mean for image location to be close together?

Does distance in neuronal activation space have a perceptual correlate?
One thing we know is that mirror images are close to each other.
We also know that mirror images are perceptually confusable.
Maybe the closer two images are the more confusable they are.

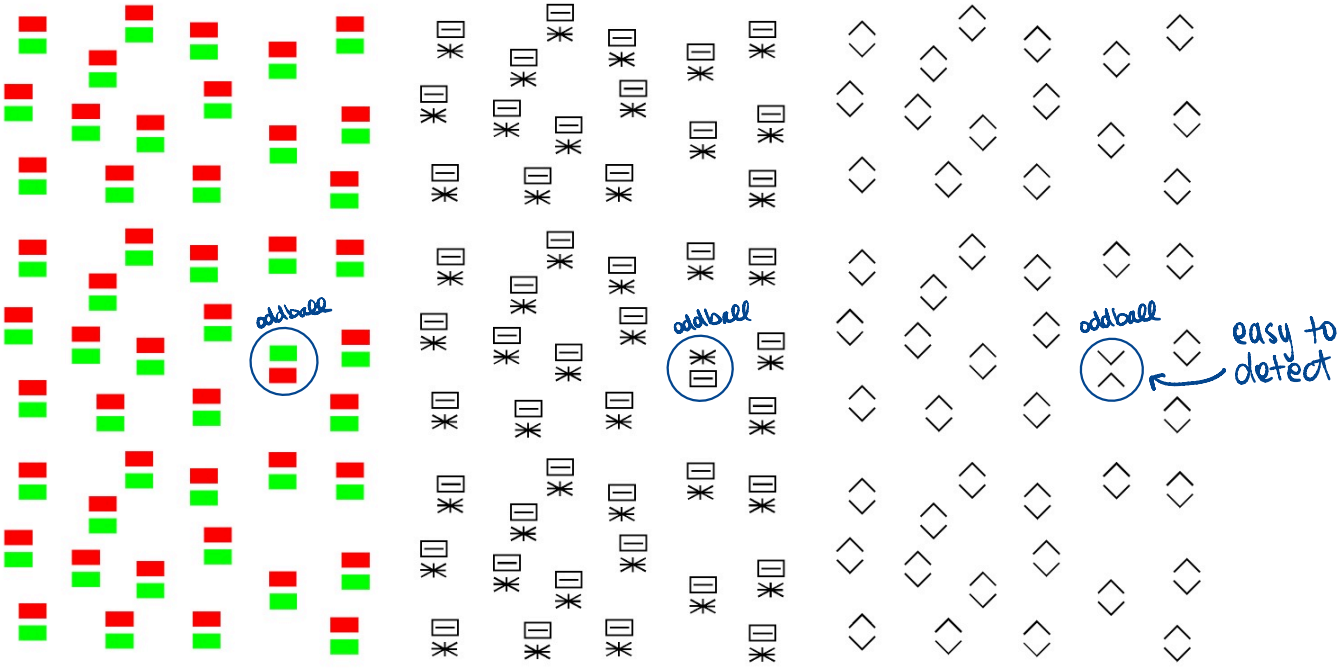
The diagram below shows the distribution of a family of stick-figure images in a plane containing the first and second principal components of a high dimensional inferotemporal neuronal activation space. Note that lateral mirror images (easily confused by humans) lie close together whereas vertical mirror images (not easily confused by humans) lie far apart.



Olson and Rollenhagen, Science 287:1506-8 (2000).

"Learing your p's & q's"

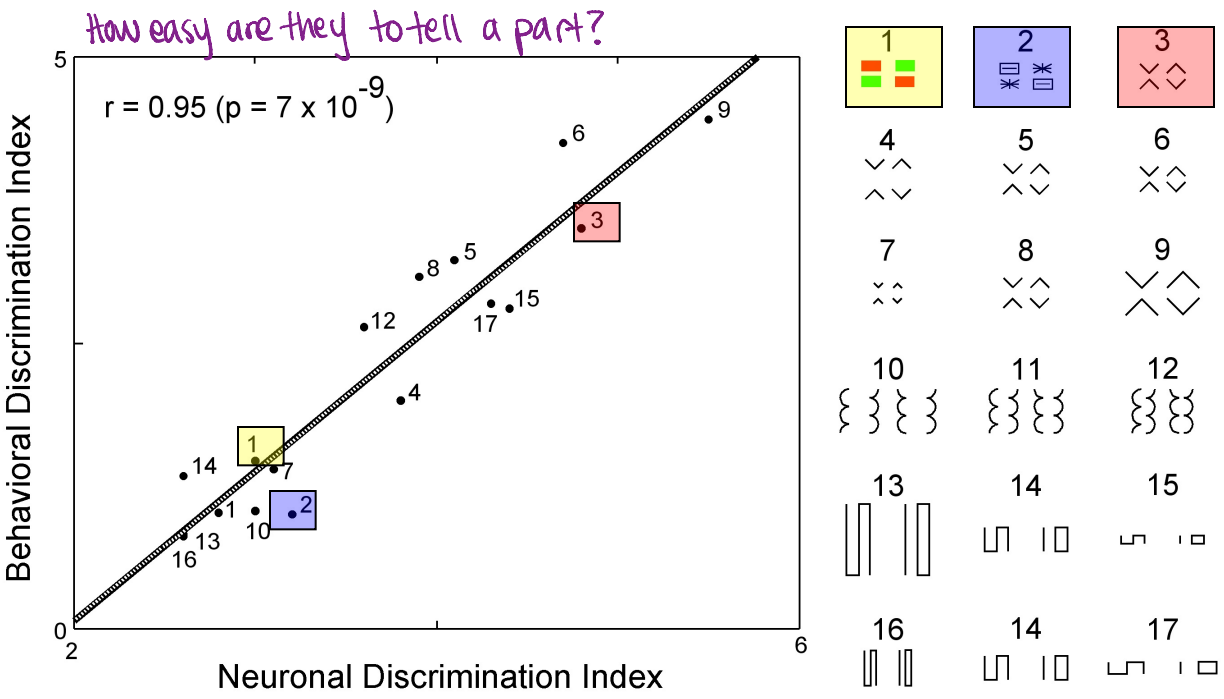
Similarity in IT Activation -> Difficulty in Oddball Search



The ability of monkey IT neurons to tell two images apart (measured as the average, across a large neuronal population, of the difference in firing rate elicited by the two images) is strongly correlated with the ability of humans to tell the images apart (measured as the speed with which they can find one embedded in a field of the others. Thus for two images to appear similar may mean no more or less than for them to elicit similar patterns of activity in IT. Note that within each array , above, there is one oddball, formed by transposition of the top and bottom elements. Only in the array on the right does the oddball pop out.

Sripati and Olson
J. Neurosci. 30:1258-1269 (2010)

Similarity in IT Activation -> Difficulty in Oddball Search



The ability of monkey IT neurons to tell two images apart (measured as the average, across a large neuronal population, of the difference in firing rate elicited by the two images) is strongly correlated with the ability of humans to tell the images apart (measured as the speed with which they can find one embedded in a field of the others. Thus for two images to appear similar may mean no more or less than for them to elicit similar patterns of activity in IT. Note that within each array , above, there is one oddball, formed by transposition of the top and bottom elements. Only in the array on the right does the oddball pop out.

Sripati and Olson
J. Neurosci. 30:1258-1269 (2010)

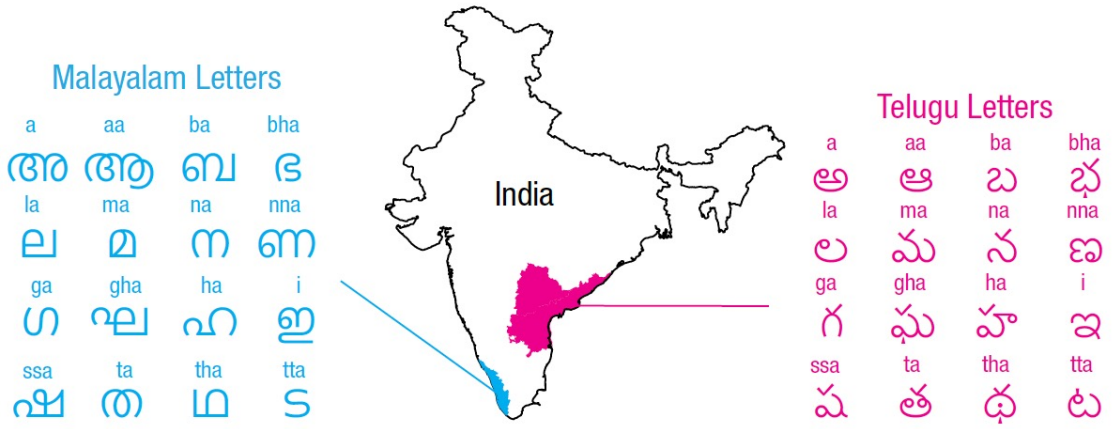
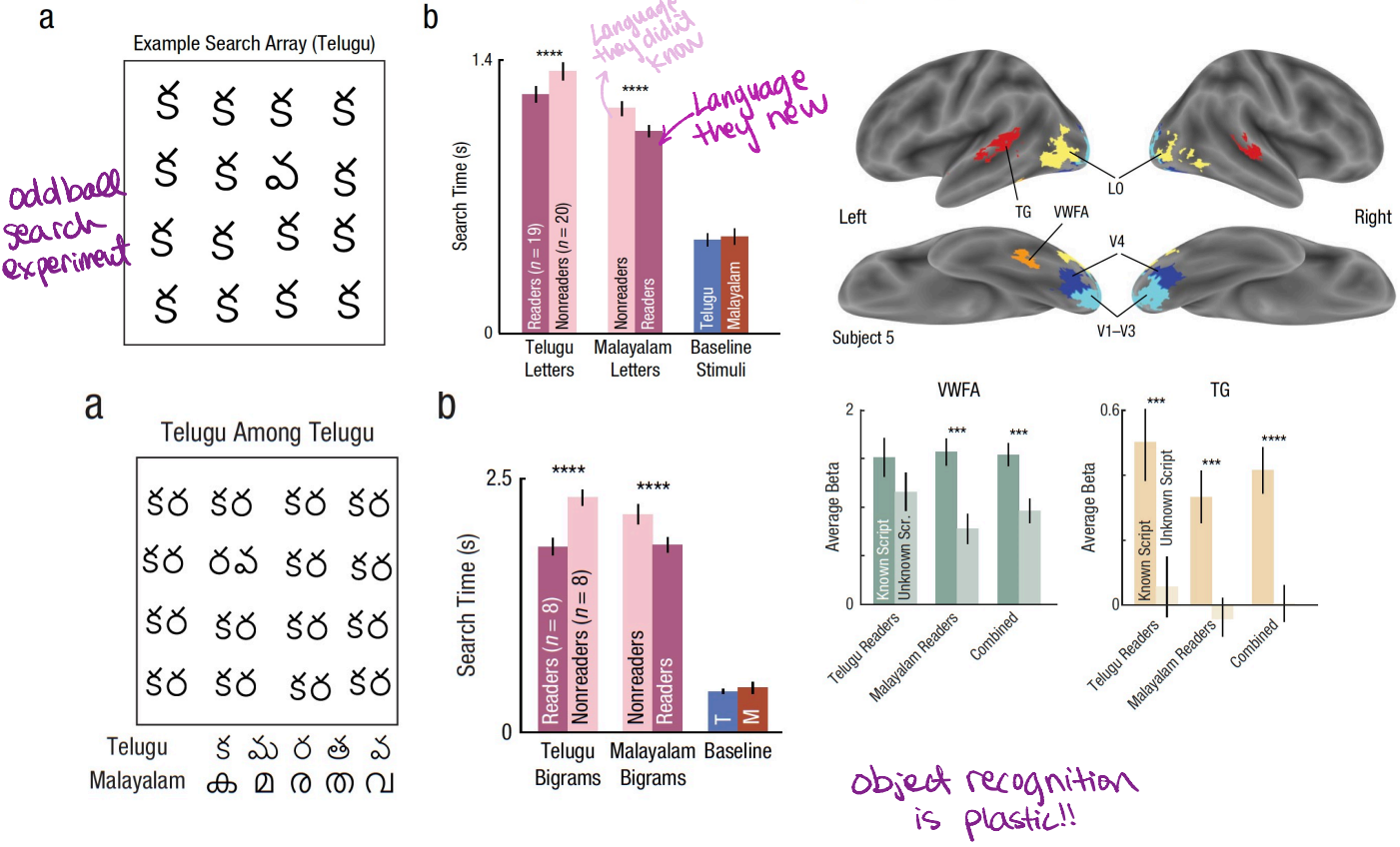


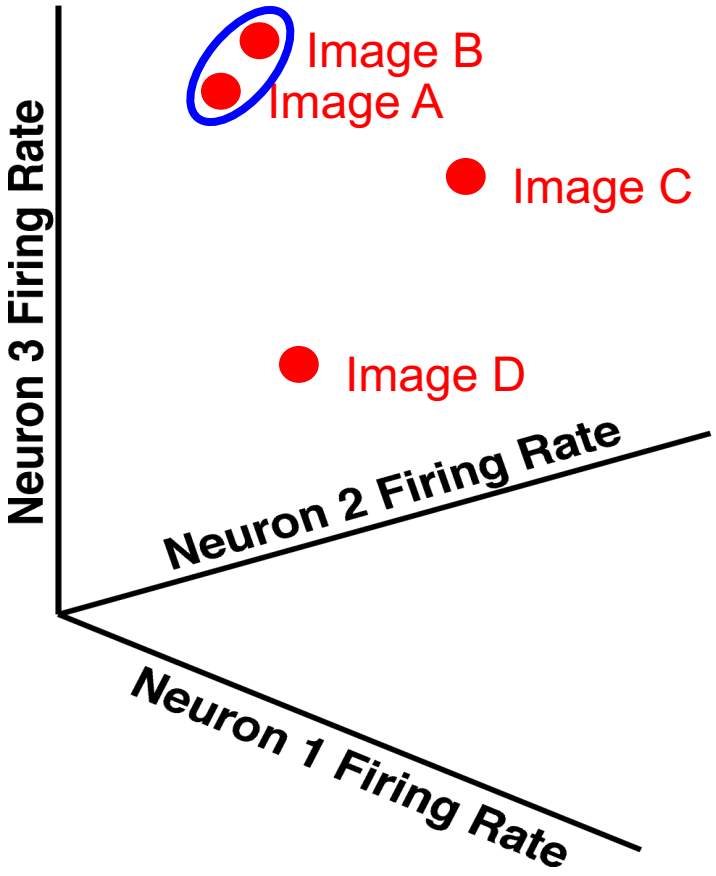
Fig. 1. Malayalam and Telugu scripts. The Malayalam and Telugu languages are spoken in geographically distinct regions in India, highlighted on the map. The scripts have distinct letter shapes but share many phonemes (indicated above each letter). Only 16 example letters are shown here from each language; Telugu has 60 letters, and Malayalam has 53 letters. The full set of stimuli is shown in Section S1 in the Supplemental Material available online. Map courtesy of Free Vector Maps (https://freevectormaps.com/).



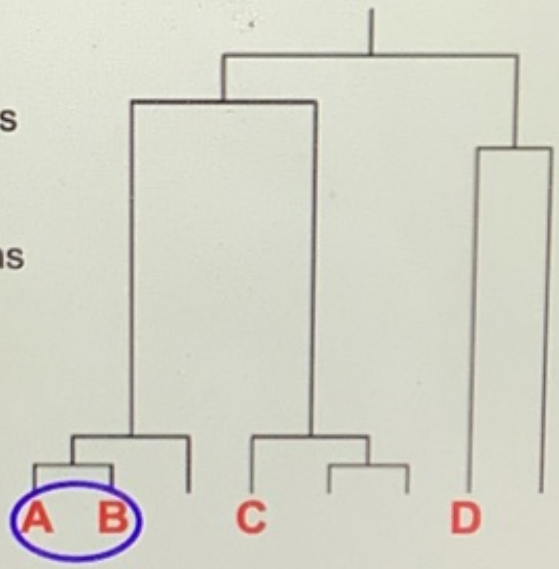
Learning a script enhances in parallel the discriminability of letters and bigrams in visual oddball search and inter-letter-distance and inter-bigram-distance in high-order visual cortex.

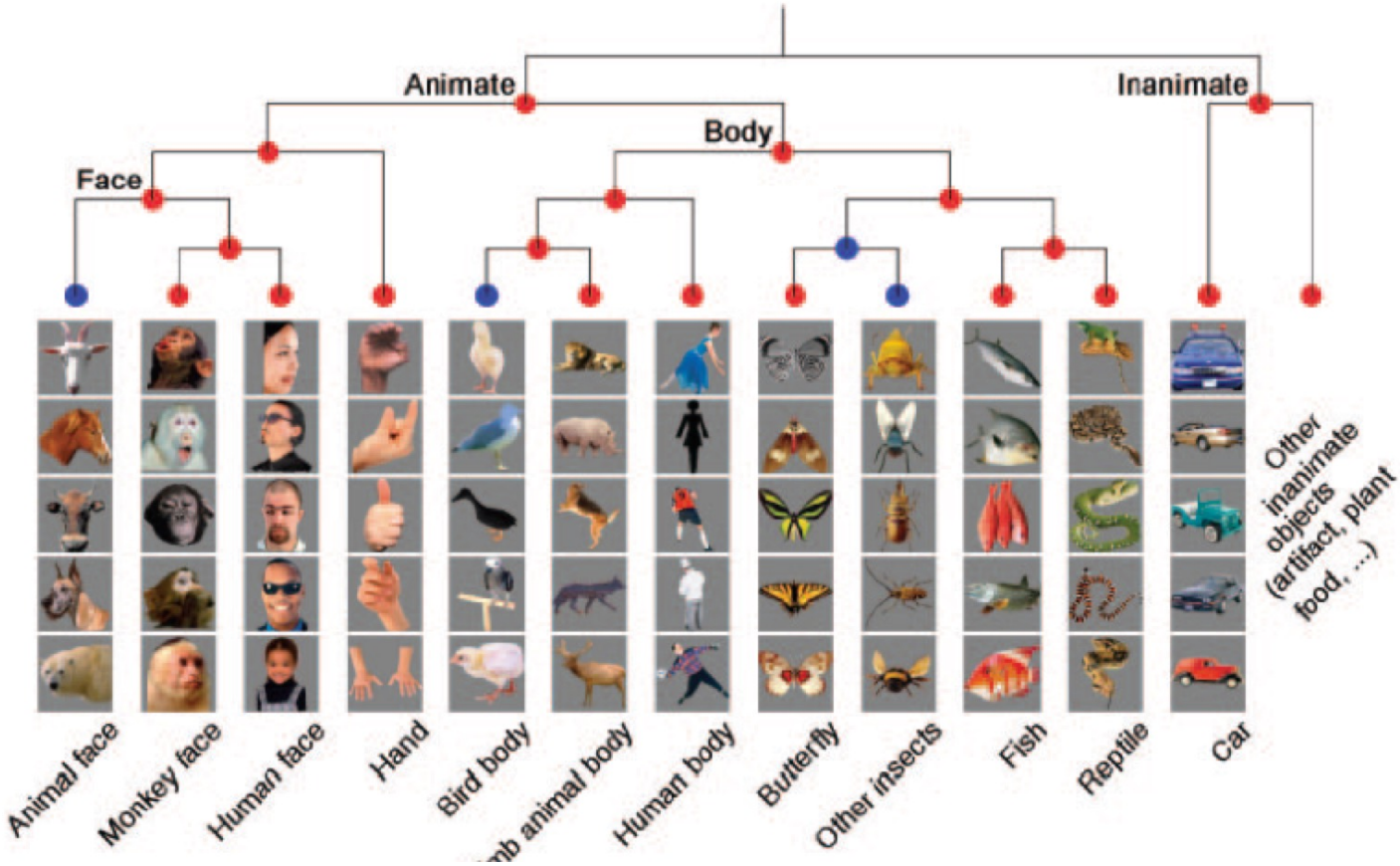
Agrawal, Hari, Arun (2019). Reading Increases the Compositionality of Visual Word Representations. Psychol Sci 30:1707-1723.

Not fixed property, but learned property



Kiani and colleagues used cluster analysis as a means for analyzing and representing inter-image distance relations in a high-dimensional activation space





Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex

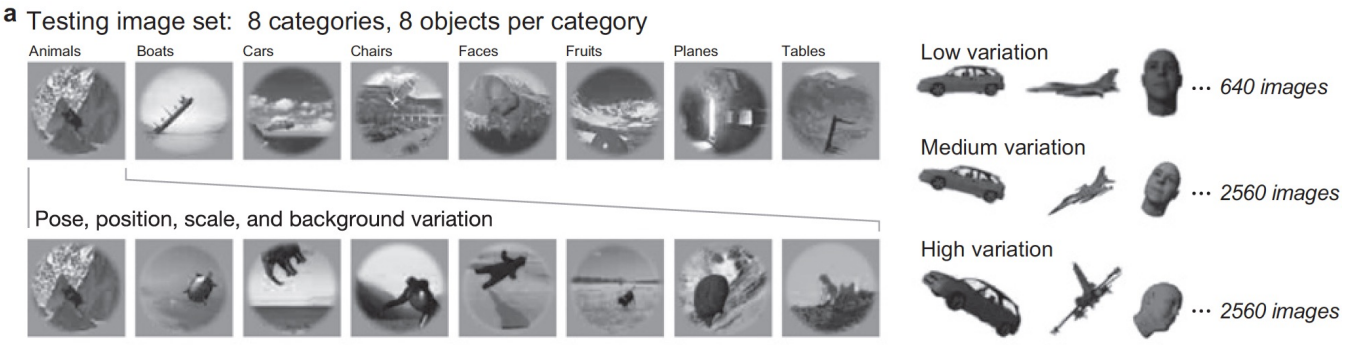
Roozbeh Kiani, Hossein Esteky, Koorosh Mirpour and Keiji Tanaka

J Neurophysiol 97:4296-4309, 2007. First published 11 April 2007; doi:10.1152/jn.00024.2007

We found that the categorical structure of objects is represented by the pattern of activity distributed over the cell population. Animate and inanimate objects created distinguishable clusters in the population code. The global category of animate objects was divided into bodies, hands, and faces. Faces were divided into primate and nonprimate faces, and the primate-face group was divided into human and monkey faces. Bodies of human, birds, and four-limb animals clustered together, whereas lower animals such as fish, reptile, and insects made another cluster. Thus the cluster analysis showed that IT population responses reconstruct a large part of our intuitive category structure.

FIG. 5. The tree reconstructed based on the neural distances. Red circles indicate the nodes significantly matching the categories. Blue circles indicate the nodes that had scores (see METHODS) significantly larger than chance score but included fewer than half of the category members. The blue nodes were added to indicate category combinations significantly matching the higher nodes. Five examples of category members are shown for each of the lowest-level categories, except for "other inanimate objects" (the rightmost node). The thirteen categories located at the lowest level are referred to as "the lowest-level categories" throughout this paper.

Images in the same category tend to be close together in inferotemporal neuronal activation space, as indicated by the results of an analysis clustering images on the basis of their proximity in activation space. It is worth thinking about whether this outcome genuinely reflects neuronal sensitivity to category membership (a learned semantic attribute) or just sensitivity to physical image attributes that are correlated with category membership.



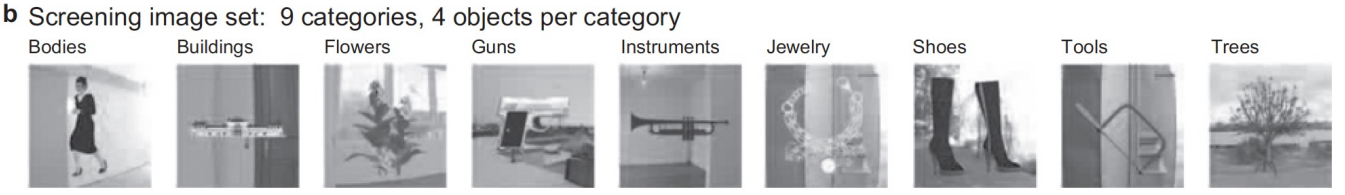
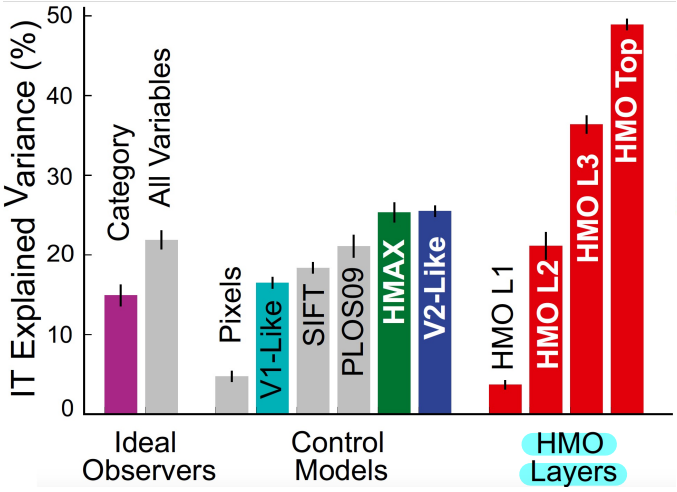


Fig. S1. (*A*) The neural representation benchmark (1) testing image set on which we collected neural data and evaluated models contained 5,760 images of 64 objects in eight categories. The image set contained three subsets, with low, medium, and high levels of object view variation. Images were placed on realistic background scenes, which were chosen randomly to be uncorrelated with object category identity. (*B*) The screening image set used to discover the HMO model contained 4,500 images of 36 objects in nine categories. As with any two uncorrelated samples of images from the world—such as those images seen during development vs. those seen in adult life—the overall natural statistics of the screening set images were intended to be roughly similar to those of the testing set, but the specific content was quite different. Thus, the objects, semantic categories, and background scenes used in screening were totally nonoverlapping with those used in the testing set. Moreoever, different camera, lighting and noise conditions, and a different rendering software package, were used.



Comparison of IT neural explained variance percentage for various models. Bar height shows median explained variance, taken over all predicted IT units. Error bars are computed over image splits. Colored bars are those shown in *A* and *B*, whereas gray bars are additional comparisons.

Yamins et al. Proc. Natl. Acad. Sci. 111: 8619-8624 (2014).

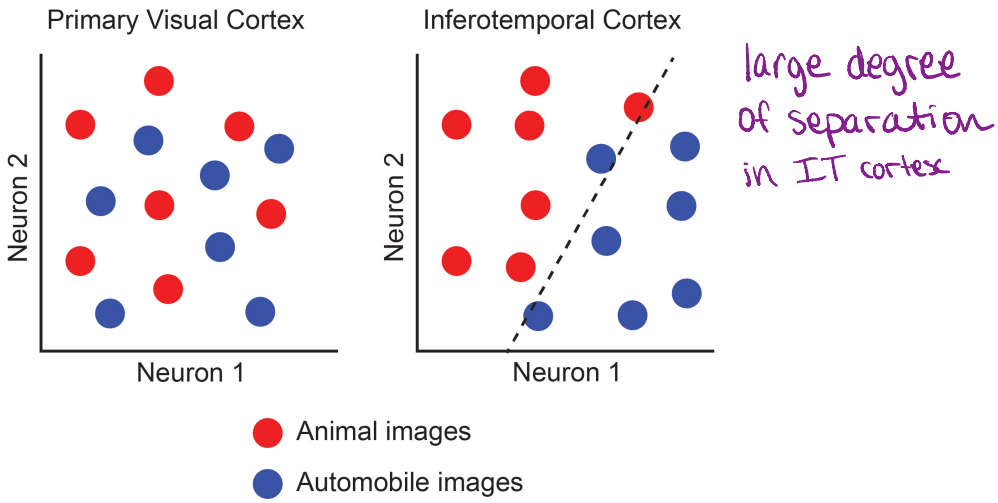
Compare representation in different models & IT neuronal space

Pairwise distances between images in macaque inferotemporal neuronal activation space are strongly correlated with pairwise distances between them in the activation space of units forming the top hidden layer of a deep convolutional neural network trained to categorize images into 8 categories regardless of pose. Network architecture was chosen by an approach based on hierarchical modular optimization (HMO) which maximized categorization performance but was agnostic with regard to the fit to inferotemporal data. Simpler models, for example ones instantiating neuronal image representations in V1 and V2, provided much poorer fits.

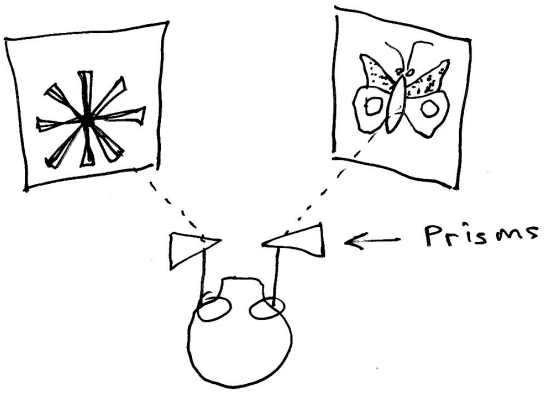
penultimate layer of deep neural network

Visual representation highly suited to make decision on semantic info

Population representations of objects become progressively more "untangled" (DiCarlo and Cox, Trends in Cognitive Neuroscience 11:333-341, 2007) as neurons at successive stages of ventral stream processing encode features that are more and more diagnostic of category membership and less and less sensitive to accidental attributes of the image arising from pose, lighting and so on. This renders images in different categories linearly separable. I.e. it is possible to identify a hyperplane in activation space that largely separates images in one category from those in the other.

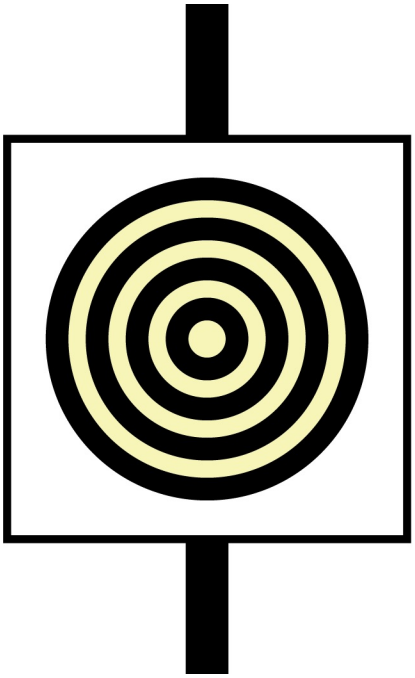


In Binocular Rivalry: Perceived = Represented in IT

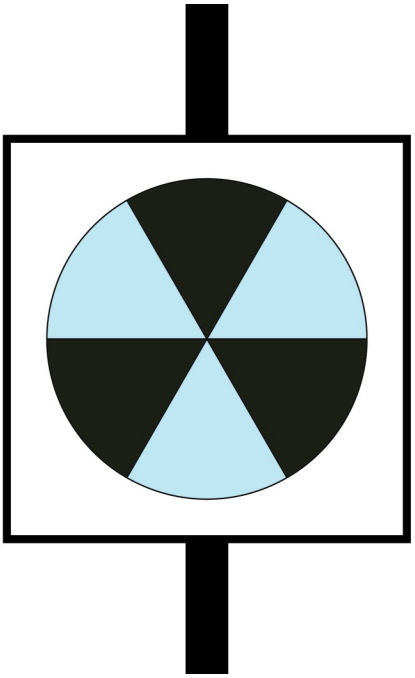


Binocular rivalry: When subjects fuse two frames, one seen through the right & one through the left eye, that contain different images, then perception will tend to alternate, with a period in the range of a fraction of a second to several seconds, between the two images.

Sheinberg & Logothetis used a paradigm in which the monkey was trained to hit one key when he saw a "sunburst" and the other key when he saw anything else. They could then select as their current "anything else" the image to which a cell in IT was responsive. Using this approach, they could ask, for example, whether a "butterfly" cell would fire when the monkey reported a butterfly but not when he reported a sunburst.



If you cross your eyes so as to fuse the two figures above, you should see the radial and concentric patterns alternating due to binocular rivalry. Focusing attention on the square helps maintain fusion.



If you cross your eyes so as to fuse the two figures above, you should see the radial and concentric patterns alternating due to binocular rivalry. Focusing attention on the square helps maintain fusion.

In Binocular Rivalry: Perceived = Represented in IT

Tricking the eyes

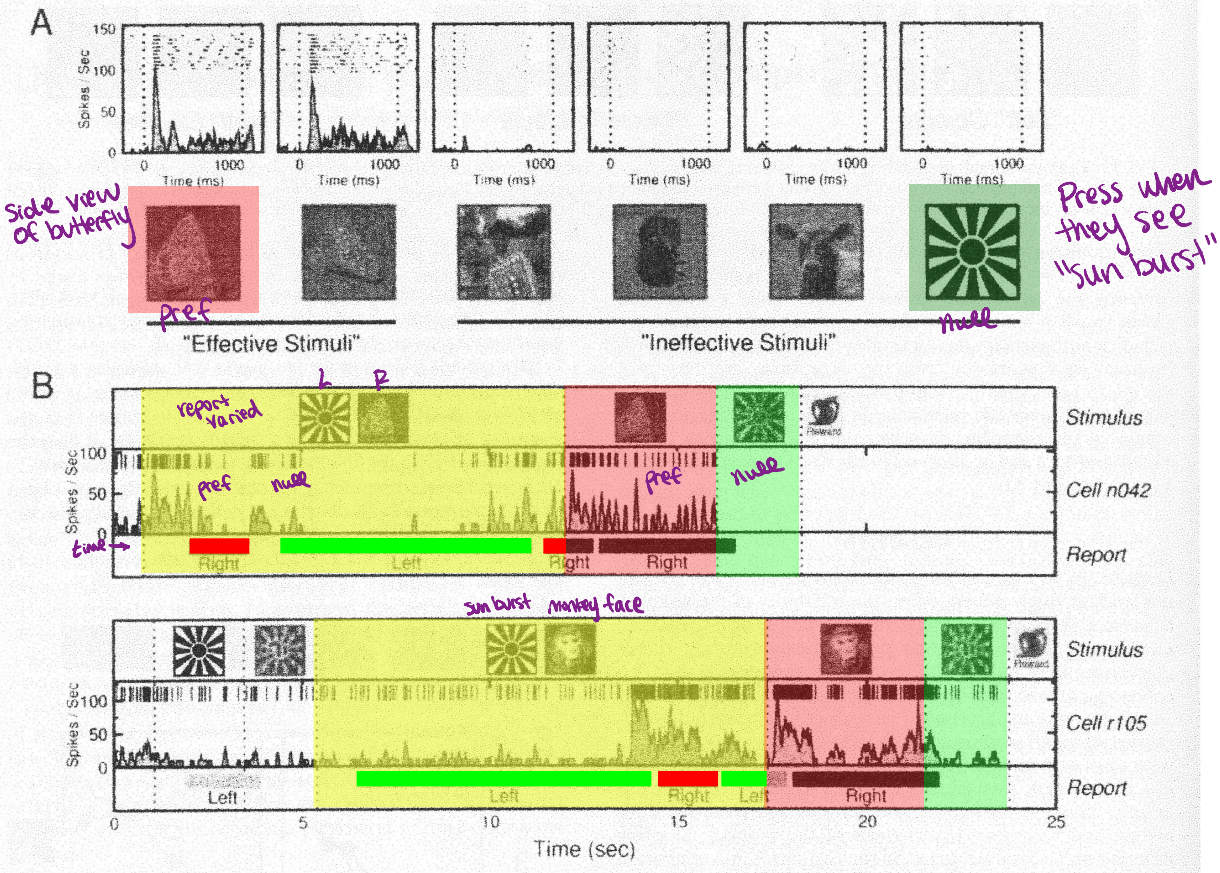


FIG. 3. Neural responses during passive viewing and during the behavioral task. (*A*) Response selectivity of an IT neuron. Effective stimuli were the two butterfly images, while almost all other tested images (30 tested, 4 shown) elicited little or no response from the cell. Each plot shows aligned rasters of spikes collected just before, during, and after the presentation of the image depicted below the graph. The smooth filled lines in each plot are the mean SDFs for all trials. The dotted vertical lines mark stimulus onset and stimulus removal. (*B*) Example observation periods taken from the behavioral task for individual cells from monkey N (*Upper*) and monkey R (*Lower*). Observation periods during behavioral testing consisted of random combinations of nonrivalrous stimuli and rivalrous periods. Dotted vertical lines mark transitions between stimulus conditions. Rivalry periods, which could occur at any time during an observation period, are shown by the filled gray background. The horizontal light and dark bars show the time periods for which the monkey reported exclusive visibility of the left-lever (sunburst) and right-lever (e.g., butterfly or monkey face) objects. Note that during rivalry the monkey reports changes in the perceived stimulus with no concomitant changes of the displayed images. Such perceptual alternations regularly followed a significant change in the neurons' activity, as shown by the individual spikes in the middle of each plot and by the SDFs below the spikes. Note the similarity of the responses elicited by the unambiguous presentation of the effective and ineffective stimuli (white regions) with those responses elicited before either stimulus becomes perceptually salient during rivalrous stimulation (gray region).

Under binocular rivalry
reported one thing or another
& IT responds accordingly

Can you predispose monkey to report a face?

Use microstimulation to impose a bias

With Electrical Stimulation: Represented in IT = Perceived

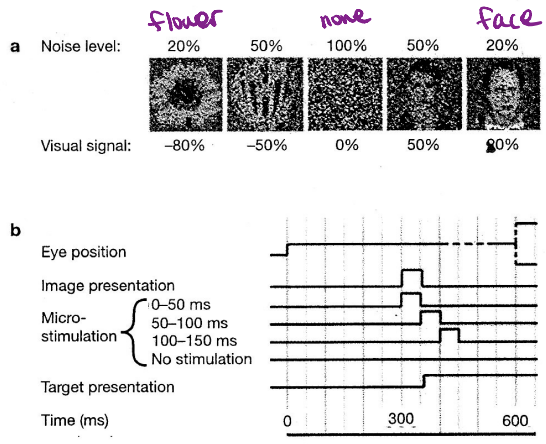
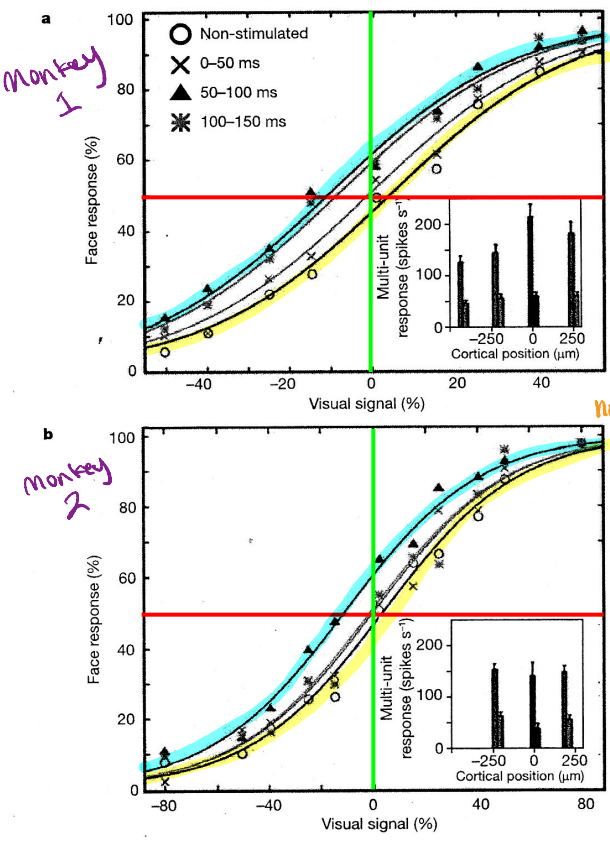


Figure 1 | Visual stimuli and event timing. In each experimental session, the neural stimulus selectivity of several neighbouring cortical sites was first determined in a fixation task using luminance-matched face and non-face greyscale images. Then, in the second part of the experiment (**a**), face and non-face images with varying amounts of noise were used in a face categorization task. **b**, Timing of events in each categorization trial. One of the four possible microstimulation conditions shown was applied randomly in each trial.



Uniform psychometric curves

50%

Stimulated

Non-Stimulated

↑ likely to report faces

50%

Figure 2 | Effect of microstimulation of two representative face-selective neural clusters in IT cortex. **a**, Monkey KH; **b**, monkey FR. Data points show the proportion of face choices for different levels of visual signal in the images for different microstimulation conditions. The curves are logistic regression fits to the data points. The insets show averaged multiunit responses of the corresponding stimulated sites and their neighbouring sites. The inset abscissa shows the cortical position of the electrode tip along the recording track relative to the stimulated site (zero on the abscissa). The inset ordinate shows the averaged multi-unit neural responses. Responses to face and non-face stimuli are represented by red and blue bars, respectively. Colours are highlighted for the stimulated site. Error bars, s.e.m.

In monkeys performing a task requiring them to report with an eye-movement (rightward or leftward) whether a foveally presented image represented a face or a non-face object, electrical microstimulation of patches of face-cells in inferotemporal cortex induced a bias to report that the image represented a face.

Afraz et al. Nature 442:692-5 2006