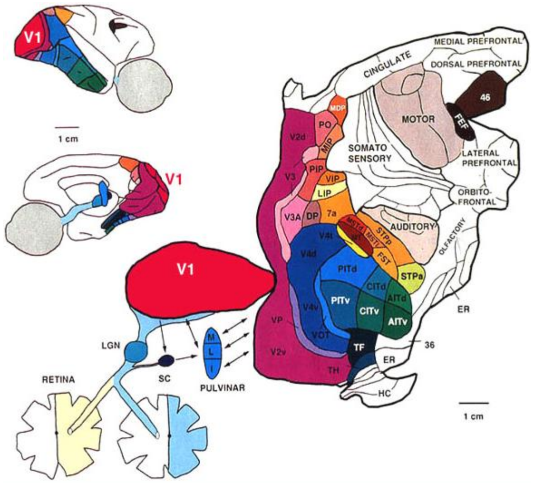
Modeling Extrastriate Areas

Dr. James A. Bednar

jbednar@inf.ed.ac.uk

http://homepages.inf.ed.ac.uk/jbednar

Higher areas



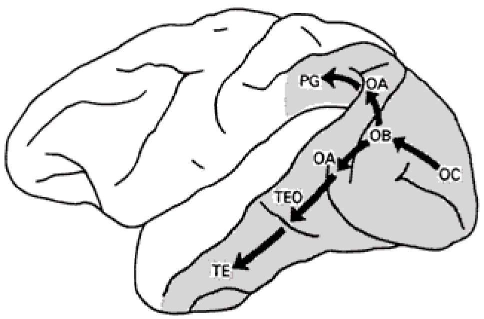
Macaque visual areas

(Van Essen et al. 1992)

- Many higher areas beyond V1
- Selective for faces, buildings, self-motion, etc.
- Not as well understood

What/Where streams

(Ungerleider & Mishkin 1982)



Typical division:

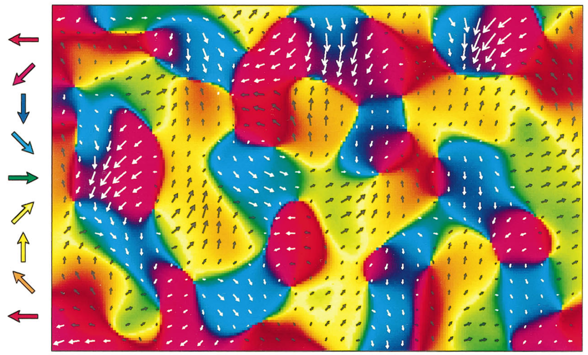
Ventral stream:

“What” pathway from V1 to temporal cortex (IT)

Dorsal stream:

“Where” pathway from V1 to parietal cortex (e.g. MT)

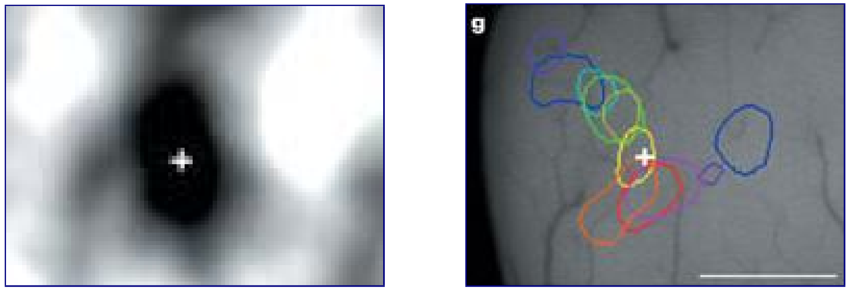
V2 OR/DR map



V2 cat direction map (Shmuel & Grinvald 1996)

- Except OD, maps found in V1 are usually also found in V2
- RFs are larger, maybe more complex (not really clear)
- Macaque V2 has complicated organization of thick/ thin/pale stripes selective for color, luminance, etc.

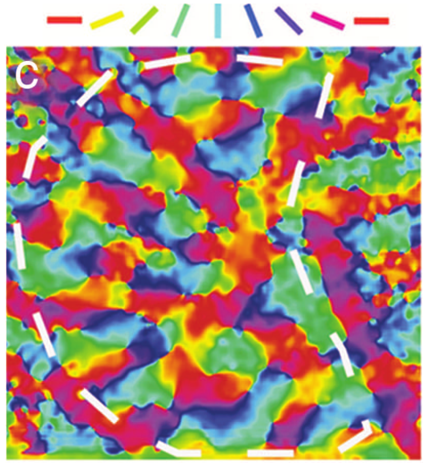
V2 Color map



Xiao et al. 2003 – Macaque; 1.4 × 1.0mm

- Like V1, color preferences organized into blobs
- Rainbow of colors per blob (Xiao et al. 2007: in V1 too?))
- Arranged in order of human perceptual color charts (CIE/DIN)
- Feeds to V4, which is also color selective

MT/V5



(Xu et al. 2006)

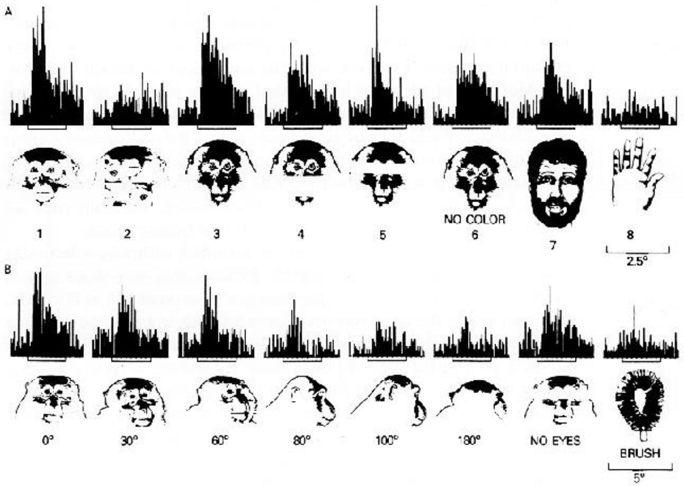
MT has orientation maps, but the neurons are more motion and direction selective

Involved in estimating optic flow

Neural responses in MT have been shown to directly reflect and determine perception of motion direction

(Britten et al. 1992; Salzman et al. 1990)

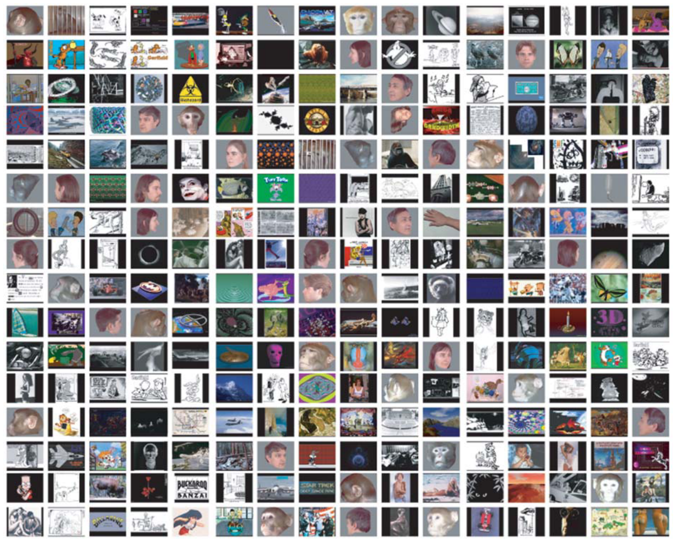
Object selectivity in IT



(Bruce et al. 1981)

Some cells show greater responses to faces than to other classes; others to hands, buildings, etc. Hard to interpret, though.

Rapid Serial Visual Presentation



(Földiák et al. 2004)

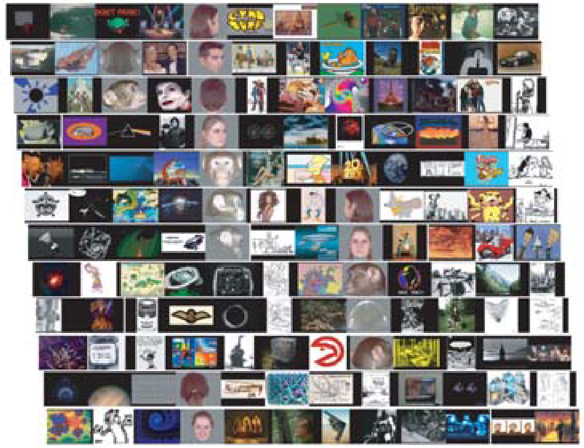
1000s of images ($> 15\%$ faces) presented to neuron for 55 or 110ms

RSVP: Face-selective neurons

(Földiák et al. 2004)

- Some monkey STSa neurons show clear preferences – top 50 faces are images
- Response low to remaining patterns
- Concern: faces are the only special category (overrepresented, aligned, blank background)

RSVP: Non-face-selective neurons

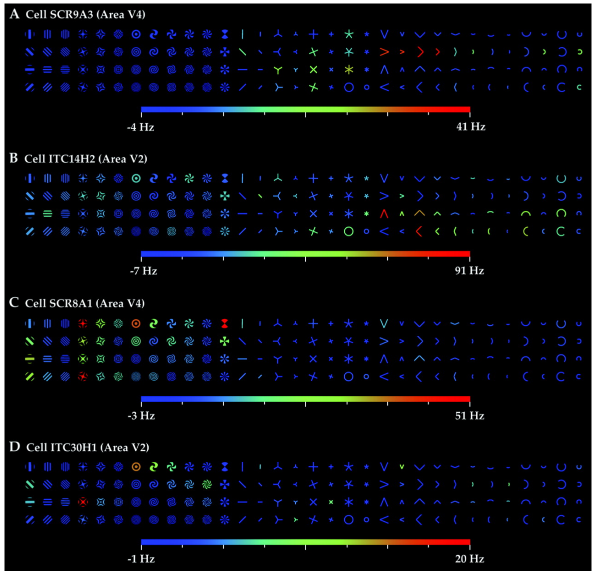


(Földiák et al. 2004)

- Other neurons don't make much sense at all
- See also Naselaris et al. (2009); mapping based on semantic category for tagged images

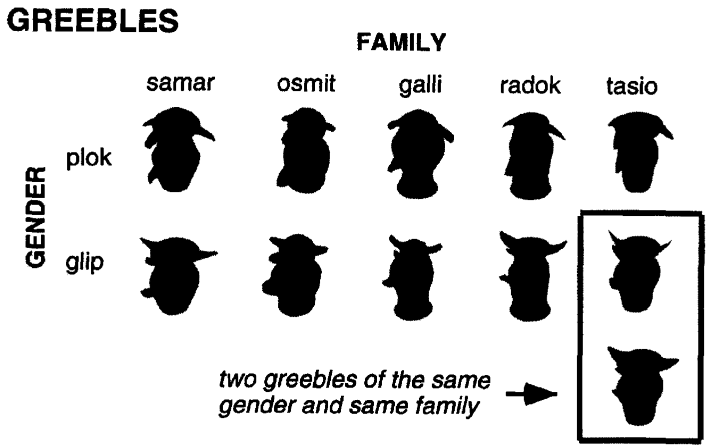
Parametric testing

Macaque; (Hegdé & Van Essen 2007)



Difficult to see differences in kind in responses to geometric stimuli across the hierarchy

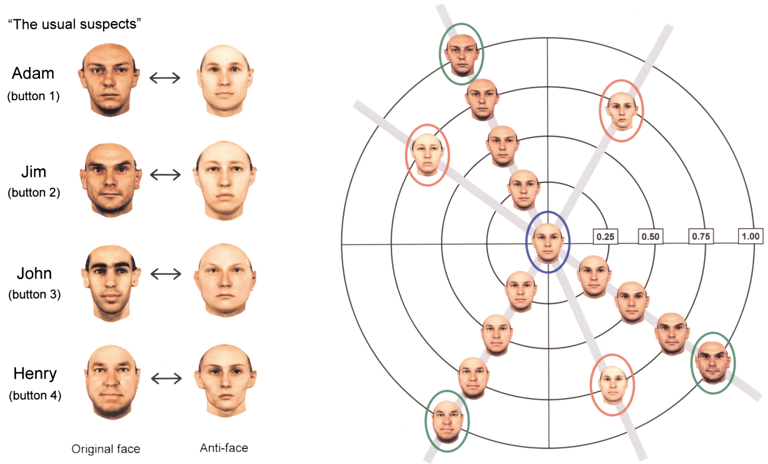
Form expertise



(Gauthier & Tarr 1997)

Most of the "specialness" of faces appears to be shared by other object categories requiring configural distinctions between similar examples.

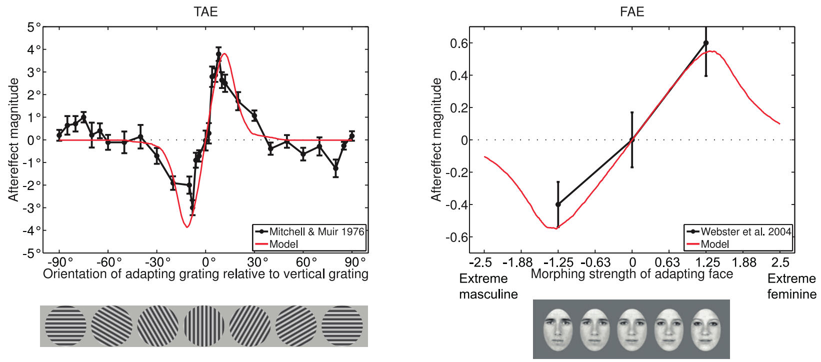
Face aftereffects



(Leopold et al. 2001)

Aftereffects are seemingly universal. E.g. face aftereffects: changes in identity judgments; blur/sharpness aftereffects, contrast aftereffects. . .

Face aftereffects same as TAE?



(Zhao et al. 2011)

People have elaborate theories about high-level aftereffects, but we successfully tested a clear prediction of the assumption that they are just like the TAE

(Zhao et al. 2011)

Invariant tuning

Higher level ventral stream cells have response properties invariant to size, viewpoint, orientation, etc.

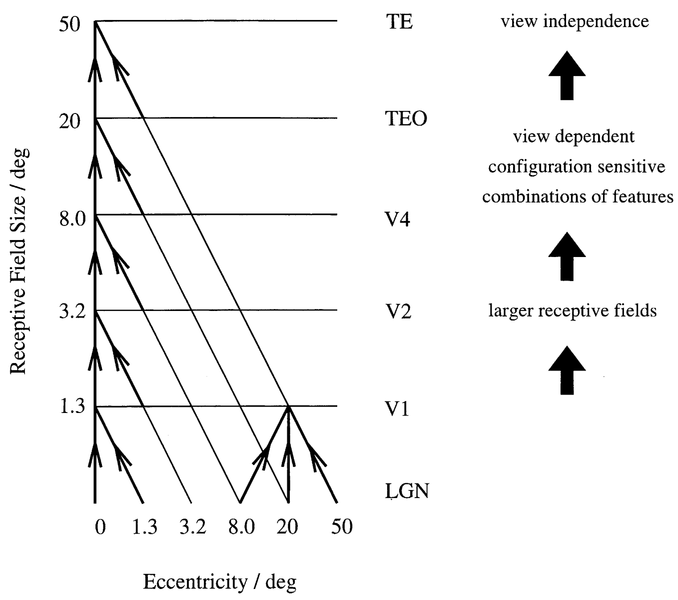
Similar to complex cells, but higher-order. E.g. can respond to face regardless of its location and across a wide range of sizes and viewpoints.

In the extreme: Jennifer Aniston neurons (Quian Quiroga et al. 2005).

Why is invariance hard?

Simple template-based models won't provide much invariance, but could build out of many such cells, or constellations of features.

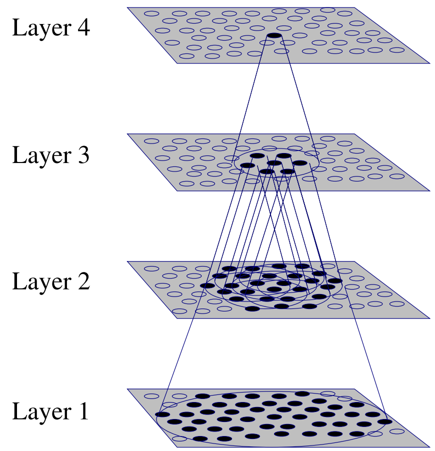
RF sizes



(Rolls 1992)

VisNet

(Wallis & Rolls 1997)



Develops neurons with invariant tuning

Assumes fixed V1 area

Ignores topographic organization

Trace learning rule

VisNet uses the trace learning rule proposed by Földiák (1991). Based on Hebbian rule for activity y^τ and input ${x_j}^\tau$:

$$\Delta w_j = \alpha y^\tau {x_j}^\tau \qquad (1)$$

but modified to use recent history ("trace") of activity:

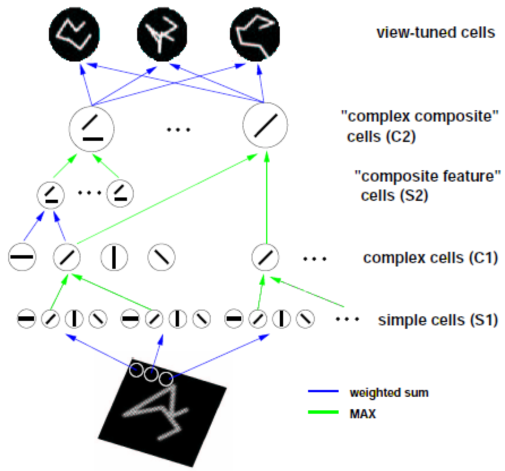
$$\Delta w_j = \alpha \bar{y}^\tau {x_j}^\tau \qquad (2)$$

$$\bar{y} = (1-\eta) y^\tau + \eta \bar{y}^{\tau-1} \qquad (3)$$

General technique for invariant responses? Need not be a special learning rule, if activity itself persists locally.

HMAX

(Riesenhuber & Poggio 1999)



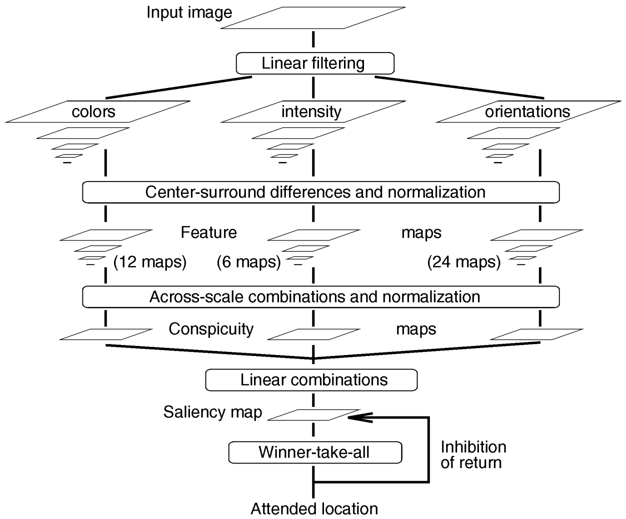
Top level (only) learns view, position, size invariant recognition

Max (C) units: nonlinear pooling, like complex cells

Linear (S) units: feature templates, like simple cells

No clear topography

Koch and Itti saliency maps



(Itti, Koch, & Niebur 1998)

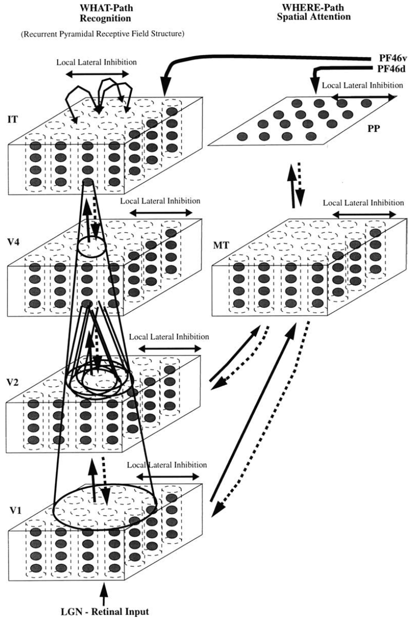
Attention model: most salient feature attended

Various feature maps pooled at different scales

Single winner: attended location

Inhibition of return: enables scanning

Other attention models



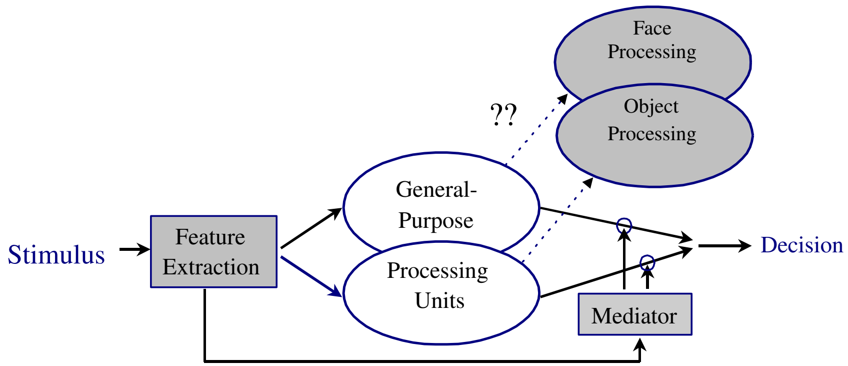
(Deco & Rolls 2004)

There are a number of other models of behavior like attention, most quite complex

Hard to tie individual model areas to specific experimental results from those areas

Also need to include superior colliculus

Modeling separate streams



(Dailey & Cottrell 1999)

Slight biases are sufficient to make one stream end up selective for faces, the other for objects

More complexities

Need to include eye movements, fovea/periphery.

At higher levels, neurons become multisensory and sensorimotor.

Eventually, realistic models will need to include auditory areas, touch areas, etc.

Output to and feedback from motor areas is also more important at higher levels.

Training data for such models will likely be harder to make than building a robot or virtual environment – will need embodied models.

Summary

- Need to include many areas besides V1
- Complexity and lack of data are serious problems
- Eventually: situated, embodied, possibly virtual, models
- May be useful to focus on species with just V1 or a few areas before trying to tackle whole visual hierarchy
- Lots of work to do

References

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, *12*, 4745–4765.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46* (2), 369–384.

Dailey, M. N., & Cottrell, G. W. (1999). Organization of face and object recognition in modular neural network models. *Neural Networks*, *12* (7), 1053–1074.

Deco, G., & Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, *44* (6), 621–642.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, *3*, 194–200.

Földiák, P., Xiao, D., Keysers, C., Edwards, R., & Perrett, D. I. (2004). Rapid serial visual presentation for the determination of neural selectivity in area STSa. *Progress in Brain Research*, *144*, 107–116.

Gauthier, I., & Tarr, M. J. (1997). Becoming a ‘Greeble’ expert: Exploring mechanisms for face recognition. *Vision Research*, *37* (12), 1673–1682.

Hegdé, J., & Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas V2 and V4. *Cerebral Cortex*, *17* (5), 1100–1116.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for

rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20* (11), 1254–1259.

Leopold, D. A., O’Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4* (1), 89–94.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63* (6), 902–915.

Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435* (7045), 1102–1107.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2* (11), 1019–1025.

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions: Biological Sciences*, *335* (1273), 11–21.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, *346*, 174–177, Erratum 346:589.

Shmuel, A., & Grinvald, A. (1996). Functional organization for direction of motion and its relationship to orientation maps in cat area 18. *The Journal of Neuroscience*, *16*, 6945–6964.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In Ingle,

D. J., Goodale, M. A., & Mansfield, R. J. W. (Eds.), *Analysis of Visual Behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, *255*, 419–423.

Wallis, G. M., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51* (2), 167–194.

Xiao, Y., Casti, A., Xiao, J., & Kaplan, E. (2007). Hue maps in primate striate cortex. *Neuroimage*, *35* (2), 771–786.

Xiao, Y., Wang, Y., & Felleman, D. J. (2003). A spatially organized representation of color in macaque cortical area V2. *Nature*, *421*, 535–539.

Xu, X., Collins, C. E., Khaytin, I., Kaas, J. H., & Casagrande, V. A. (2006). Unequal representation of cardinal vs. oblique orientations in the middle temporal visual area. *Proceedings of the National Academy of Sciences of the USA*, *103* (46), 17490–17495.

Zhao, C., Seriès, P., Hancock, P. J. B., & Bednar, J. A. (2011). Similar neural adaptation mechanisms underlying face gender and tilt aftereffects. *Vision Research*, *51* (18), 2021–2030.