



ELSEVIER

Transformation of shape information in the ventral pathway

Charles E Connor, Scott L Brincat and Anitha Pasupathy

Object perception seems effortless to us, but it depends on intensive neural processing across multiple stages in ventral pathway visual cortex. Shape information at the retinal level is hopelessly complex, variable and implicit. The ventral pathway must somehow transform retinal signals into much more compact, stable and explicit representations of object shape. Recent findings highlight key aspects of this transformation: higher-order contour derivatives, structural representation in object-based coordinates, composite shape tuning dimensions, and long-term storage of object knowledge. These coding principles could help to explain our remarkable ability to perceive, distinguish, remember and understand a virtual infinity of objects.

Addresses

Krieger Mind/Brain Institute and Department of Neuroscience, Johns Hopkins University, Baltimore, MD 2128, USA

Corresponding author: Connor, Charles E (connor@jhu.edu)

Current Opinion in Neurobiology 2007, 17:140–147

This review comes from a themed issue on
Cognitive neuroscience
Edited by Keiji Tanaka and Takeo Watanabe

Available online 21st March 2007

0959-4388/\$ – see front matter

© 2007 Elsevier Ltd. All rights reserved.

DOI [10.1016/j.conb.2007.03.002](https://doi.org/10.1016/j.conb.2007.03.002)

Introduction

The world is familiar and comprehensible to us because we recognize and understand the objects it contains. We identify, distinguish and evaluate objects based on their shapes, which range from simple (letters and numbers) to extremely complex (faces). This is one of the most computationally daunting tasks the brain performs, owing to the complexity and variability of the input data (retinal images of objects) and the high dimensionality of object shape. It seems trivial or even transparent to us only because one of the two major pathways in visual cortex [1] is dedicated to continuously processing object information with extraordinary accuracy and rapidity [2••]. This object-processing pathway (also known as the ventral, temporal or ‘what’ pathway) transforms retinal signals into object representations that are explicit enough to support our vivid appreciation of object structure, compact enough to be stored in memory, and stable enough to generalize across different viewing conditions (Figure 1). The neural algorithms that make this possible are not yet understood, and it has proven difficult to duplicate human

visual performance using computer vision systems at the present stage.

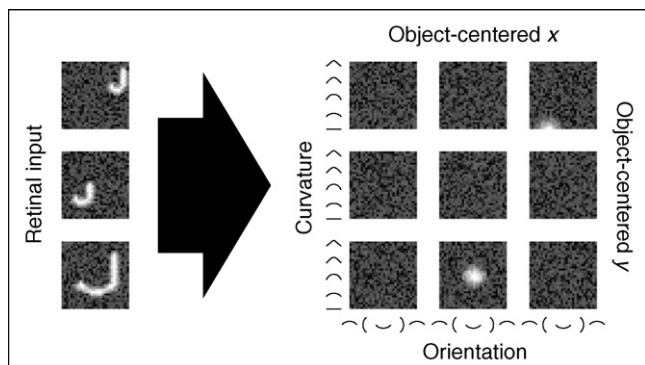
The ventral pathway runs from primary visual cortex (V1) and secondary visual cortex (V2) through area V4 and then into a series of further processing stages in ventral occipitotemporal cortex. In humans, these further stages include area V8 [3] (alternatively labeled V4 by some authors), the lateral occipital complex [4], and parts of the fusiform and parahippocampal gyri [5,6]. The general functionality of these areas can be studied using functional magnetic resonance imaging (fMRI), which has revealed, for example, regions specialized for face, body and scene processing [6–8]. Algorithmic-level shape processing can be studied using electrode recordings in monkeys, which have similar visual capacities and a highly analogous organization of ventral visual cortex (V1, V2, V4 and multiple stages in inferotemporal cortex [IT], including face and body patches) [9]. Here, we review human and monkey experiments from the past two years that shed light on how the ventral pathway transforms retinal signals into useful object representations.

Higher-order contour derivatives

It is well-established that the first stage in the ventral pathway transformation involves the extraction of local orientation and spatial frequency information [10]. Orientation is a first-order derivative that efficiently encodes elongated contrast regions that correspond to object contours. As a result, the transformation that occurs in V1 maximizes sparseness (minimizes the number of active neurons) in the representation of natural images [11,12]. This constitutes a major step towards more compact representation of useful object information.

At subsequent processing stages in area V2, area V4 and posterior IT, neurons process progressively larger image regions. On these larger scales, contour orientation is more likely to change, either gradually (broad curves) or abruptly (sharp curves or angles). There is now substantial evidence that ventral pathway neurons explicitly encode these orientation changes in terms of higher-order contour derivatives. Earlier experiments [13] showed that V4 neurons can exhibit combined selectivity for orientation and curvature (rate of change in orientation, a second-order derivative; V4 neurons can even be sensitive to spirality, a third-order derivative [14]). For example, a given V4 neuron might respond most strongly to curved convex boundaries that project toward the right. Recent experiments indicate that curvature processing begins by the V2 level [15,16] and that curvature representation is prominent across multiple higher stages in IT [17,18]. As

Figure 1



Transformation in the ventral pathway. The initial representation at the retinal level of even a simple shape such as the letter 'J' is distributed across many neurons. The shape information in this representation is implicit (i.e. difficult to read or decode): it would take many computational steps to determine that any given set of input values on these retina-like 30×30 arrays (black/gray/white pixel pattern, left) corresponds to a J. (Just looking at the pattern is cheating — you are using the computational power of your own ventral visual pathway to do the decoding.) Finally, the representation of a J at this level is extremely variable: the set of active neurons changes completely depending on size and position of the letter. Here, three different views are shown (left). The ventral pathway must transform this abstruse representation into one that is much more compact, explicit and stable. This figure schematizes a transformation into a simple structural representation, in which the J would be represented by just two activity peaks, one corresponding to the vertical straight line at the top right and the other corresponding to the curved horizontal line at the bottom. (The 3×3 grid of panels represents object-centered position; curvature and orientation are plotted recursively within this grid.) This review article discusses recent evidence for structural coding and other transformation processes that could lead to object representations that are compact, explicit and stable enough to support our remarkable perceptual capacities.

in V4, inferotemporal neurons exhibit combined selectivity for curvature and orientation of contour fragments, curvature tuning is signed (i.e. cells respond differentially to convex and concave contours), and the representation is biased towards higher curvature values [18,19].

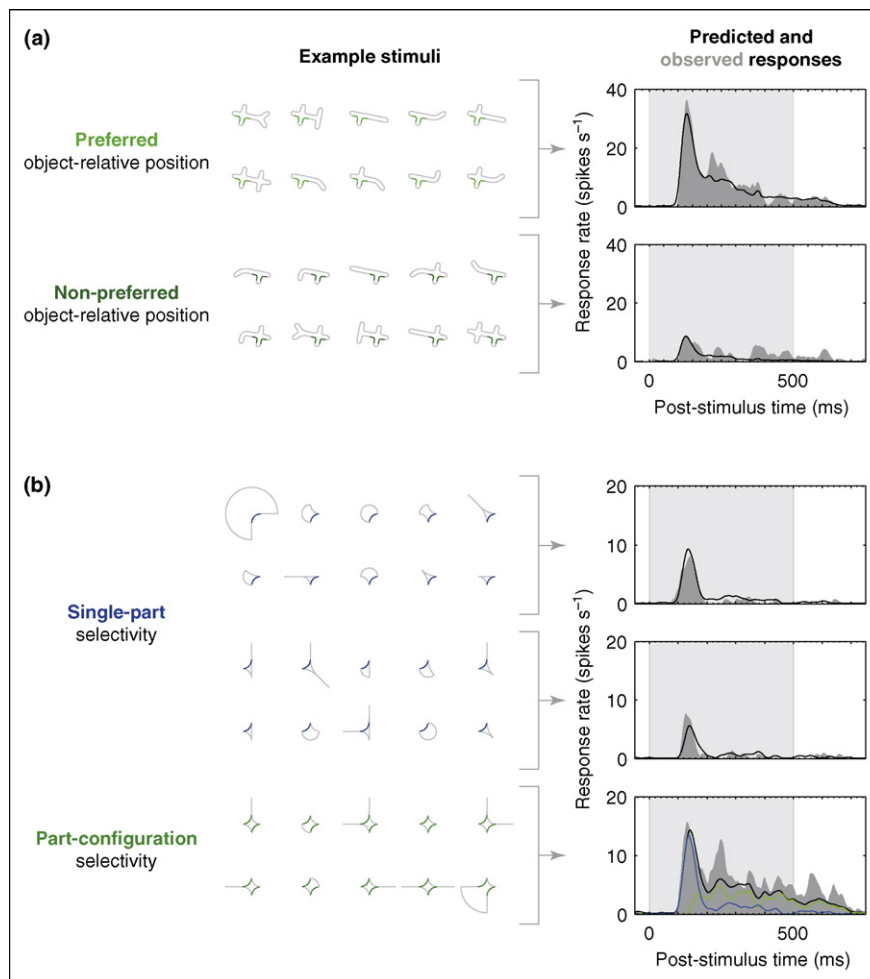
Because curvature signals efficiently encode even larger contour fragments that encompass multiple orientations, they represent a further step towards compact object representation. They also explicitly encode a fundamental aspect of shape that has strong perceptual valence for human observers [20–23]. For example, a recent study demonstrates that, contrary to standard theory, texture boundaries do not depend exclusively on orientation discontinuities: curvature discontinuities alone produce striking texture boundaries, reflecting the strong representation of curvature at the neural level [24]. It has also recently been shown that similar curvatures in close spatial proximity interfere perceptually [25], consistent with the idea of coarse basis-function coding of curvature at the neural population level [26].

Structural representation in object-based coordinates

Many theories of shape processing [27,28] are based on the idea of structural representation — that is, shape description in terms of object parts and their positional and connective relationships. Structural codes are compact, because even complex shapes comprise a manageable number of parts. Structural codes are highly generative, because even a limited basis set of different elements can be combined in so many ways. Thus, a finite number of neurons that encode object parts can represent a virtual infinity of object shapes, in the same way that just 26 letters of the alphabet can represent millions of words. Explicit representation of object structure would also explain our immediate cognitive access to information about parts and their relationships: if asked to describe an object, we usually talk about its parts and how they are put together. Finally, structural representations could be more stable across changes in viewing conditions. A key tenet of most structural theories is that part relationships are represented in a spatial reference frame centered on the object itself. This would require a major transformation of position information from the original retinotopic reference frame, but it would confer stability across changes in object position. Additionally, stability across viewing distance could be achieved if the reference frame scaled with object size, and stability across viewing angle could be achieved if the reference frame rotated with some definable axis in the object itself.

Electrode recording experiments in the monkey ventral pathway have yielded evidence of structural representation in an object-centered reference frame. V4 neurons tuned for contour fragment orientation and curvature are also strongly sensitive to the object-relative position of contour fragments [19]. For example, a given V4 neuron might respond to objects that have convex curvature at the bottom right (e.g. a lowercase 'b') but not to those that have convex curvature at the top right (e.g. a lowercase 'p'). Structural coding in an object-based reference frame is even more prominent at the next processing stage in posterior IT, where neurons integrate information about multiple contour fragments (typically 2–4) [17]. The example cell in Figure 2a is sensitive to concavities oriented towards the lower right and concavities oriented towards the lower left. (This was determined by fitting response functions across a large number of stimuli not shown here.) As in V4, the same local contour information evokes strong responses in a variety of global shape contexts (top rows, Figure 2a). Also as in V4, responses are acutely sensitive to object-relative position: the same configuration of two concavities evokes little or no response when it appears to the right of object center rather than to the left (Figure 2a, bottom rows). In more anterior parts of IT, sensitivity to object-relative position seems to be organized at the columnar level [29]. Most

Figure 2

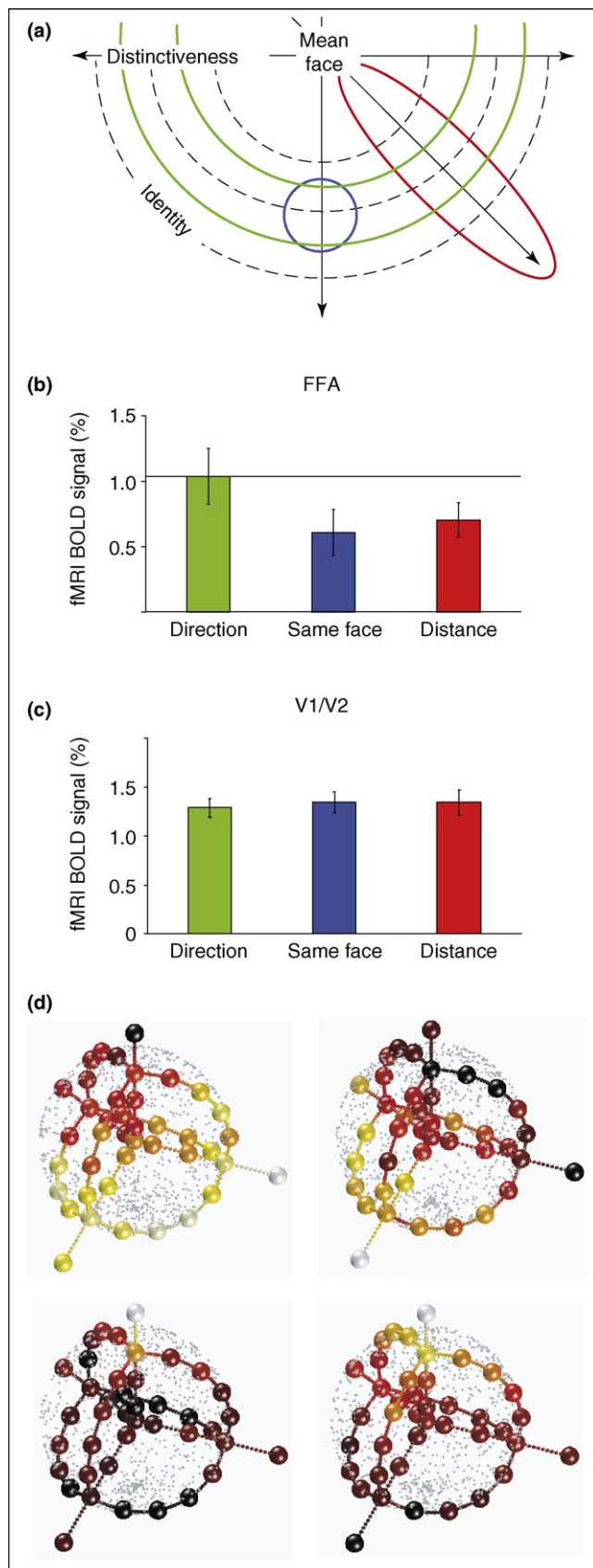


Structural representation of shape in object-centered coordinates. **(a)** Tuning for object-centered position of contour fragments. This posterior inferotemporal cell was tested with a set of 1882 stimuli comprising different combinations of contour elements that varied in curvature, orientation and relative position (for details, see [17]). The best-fitting response function (which yielded a correlation of 0.89 between observed and predicted response rates) revealed strong selectivity for a combination of concave contour fragments oriented towards the lower right and towards the lower left (highlighted in green). The cell was highly sensitive to the position of these contour elements relative to the object. It responded strongly to concavities at the left of object center (bright green), as exemplified by the average response (gray histogram) to the stimuli in the top two rows. (The black curve is the response profile predicted by the best-fitting model; the light gray shading indicates the 500 ms stimulus-presentation period.) The cell responded only weakly to the same concavities at the right (dark green; bottom two rows). This tuning for object-relative position was consistent across changes in absolute position of the stimuli. Object-centered position coding is crucial for stable structural representation. **(b)** Temporal evolution of multipart configuration signals. At early time points after stimulus onset (near 100 ms), this posterior inferotemporal cell responded to stimuli that contained lower-left oriented concavities (blue segments, top two rows), upper-left oriented concavities (blue segments, middle two rows) or both (green segments, bottom two rows). Thus, the shape information conveyed by the cell during this period was relatively ambiguous. By 200 ms after stimulus onset, the cell was responsive only to stimuli containing both contour elements, and thus provided an explicit signal for a configuration of two apposed concavities defining a narrow, diagonally-oriented neck. The blue and green curves show the linear and nonlinear factors that predicted these two phases. Similarly, across the population, ambiguous linear signals dominate early responses and explicit nonlinear signals emerge over the course of ~60 ms, presumably through network processes that refine the representation of multipart configurations (for details, see [31]).

strikingly, neurons in the superior temporal sulcus that are involved in representing body motion seem to operate in a reference frame not only centered but also oriented with respect to the body, because the same motion can evoke responses across different body orientations [30[•]].

At the V4 level, where most neurons represent individual parts (contour fragments) and their object-centered positions, configurational relationships between parts are implicit in the population response [26] — that is, they could be inferred by comparing object-relative positions. In IT, integration across part signals could generate

Figure 3



explicit representations of multipart configurations (with a concomitant gain in compactness). However, linear summation of part signals would be inherently ambiguous, because a given response level could correspond to a variety of parts or part combinations. In fact, response functions in posterior IT range from primarily linear to primarily nonlinear, and there is a strong trend towards greater nonlinearity across time following stimulus onset [31]. The example cell in Figure 2b is sensitive to concavities oriented towards the lower right and concavities oriented towards the upper left. At early time points, ~100 ms following stimulus onset, this cell responds to stimuli containing just one part or the other (upper rows) in addition to stimuli containing both parts (bottom rows). This early response phase can be accurately modeled as a linear summation across two tuning regions in the contour fragment domain (blue curve, bottom histogram). By ~200 ms, however, responses to individual parts drop out, and the response function of the neuron resolves into an explicit signal for the multipart configuration (of two apposed concavities defining a narrow diagonal neck). This later phase is captured by a nonlinear product term based on the two tuning regions (green curve). Likewise, across the neural population, linear summation dominates at early time points, and nonlinear selectivity for multipart configurations evolves gradually, peaking near 200 ms after stimulus onset. The delayed emergence of explicit signals for multipart configurations could explain why shape judgments that depend on structural nuances

Evidence for a polar domain centered on the average face. **(a)** Schematic representation of neural face space, with the mean or average face at the center. Along radial directions (red ellipse), faces that have the same basic identity or set of geometric characteristics would become increasingly distinct or caricature-like. Along circumferential directions (green annulus), face identity would change. **(b)** fMRI cross-adaptation experiment in the fusiform face area (FFA). Successive presentations of a single face stimulus (blue circle in a) evoke low activity owing to response adaptation (blue bar). Successive presentations of faces that differ in identity evoke higher activity because they stimulate different neural populations (green bar). Successive presentations of faces that have the same identity but different distinctiveness evoke lower activity (red bar). This cross-adaptation between stimuli indicates that the same neural population is being stimulated. Other experiments in the same study show that overall activity increases with distance from the average face. Together, these results suggest that any given FFA neuron responds along one specific direction radiating outwards from the average face, and that its response level encodes distinctiveness along that direction. **(c)** The lack of a similar cross-adaptation difference in early visual cortex (V1/V2) shows that the FFA results are not explained by low-level stimulus factors (for further details, see [45]). **(d)** An analogous experiment on individual neurons in monkey inferotemporal cortex (IT). Each of the four subplots represents the responses of one neuron. The colored balls indicate response rates for the average face (center), for stimuli along radial morph lines leading outwards towards four distinct faces, and for stimuli along circumferential morph lines between the distinct faces. Dark red balls correspond to low response rates, and red, orange, yellow and white correspond to increasingly higher response rates (for details, see [46]). Panels (a–c) are reproduced, with permission, from [45]; (d) is reproduced, with permission, from [46].

(e.g. recognizing individual faces) take longer than other shape judgments [32,33].

Composite tuning dimensions for face representation

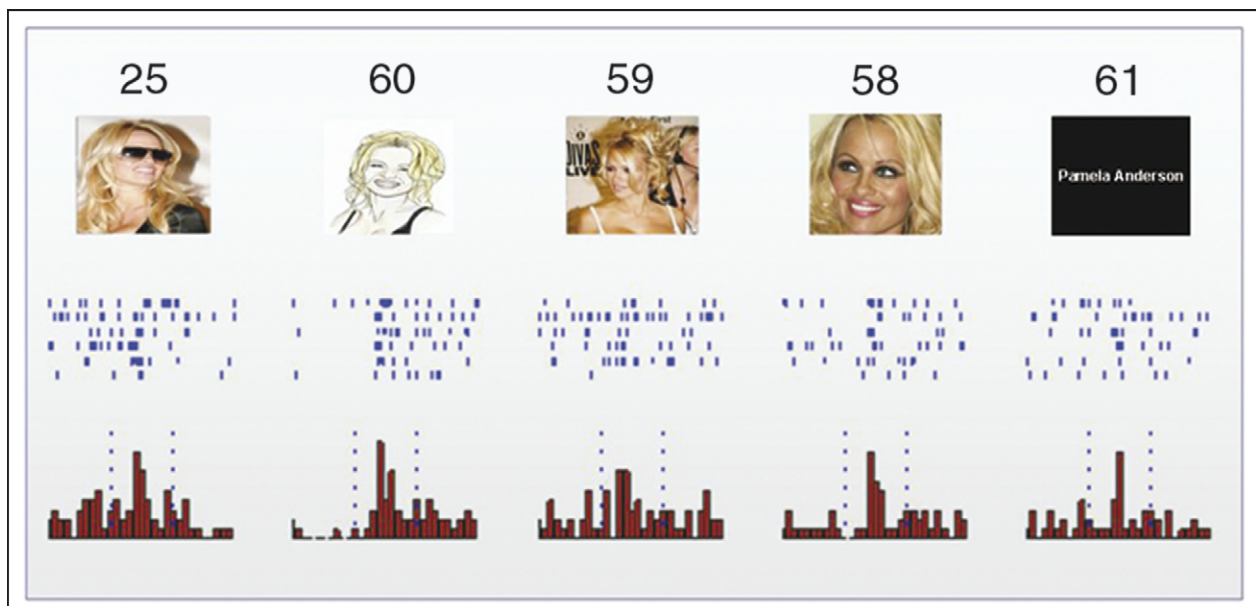
Further integration can produce composite tuning dimensions that summarize large amounts of geometric detail. For example, our exquisite expertise in discriminating faces must depend on high-level neurons that are sensitive to complex combinations of simpler structural parameters [34]. Face perception is so specialized that face-selective regions of cortex can be identified at the gross anatomical level. These regions seem to be truly specialized for faces, and not just for general discriminatory expertise [35] or configurational processing [36]. They contain an overwhelming preponderance of face-selective neurons [37**], and local electrical stimulation of these areas evokes face percepts [38**]. (However, not surprisingly, there is evidence that this powerful neural machinery can be co-opted for other complex shape-processing tasks [39].)

If geometric tuning dimensions are sufficiently complex, we no longer have immediate cognitive access to the underlying structural information. In describing faces, we do not report exact distances between the eyes, nose and mouth. Instead, we describe faces by category (e.g.

‘feminine’ and ‘Caucasian’) or by comparison (e.g. ‘similar to Pamela Anderson’). Such descriptions must reflect highly composite dimensions developed through exposure to many exemplars. The ‘morph lines’ between exemplars would become the tuning dimensions in which neurons encode identity [40]. Psychophysical experiments demonstrate that this dimensionality can be manipulated by controlling exposure to exemplars [41**].

One idea that has gained recent support is that face-processing dimensions are defined relative to an ‘average’ face, learned through statistical integration across the huge sample of faces we encounter during life. This might explain the effectiveness of caricature in conveying identity [42,43]. Psychophysical experiments have shown that face adaptation biases perception in the direction from the adapting face to the average face [44]. Consistent with this, activity in the fusiform face area (FFA) increases with distance (distinctiveness) from the average face but not with distance from a non-average face [45]. In addition, FFA responses cross-adapt for stimuli that lie along the same direction from the average face but not for stimuli that lie along tangential directions (Figure 3). These fMRI results and corresponding data at the single-cell level [46] suggest a polar organization of face dimensionality centered on the average face.

Figure 4



Explicit representation of learned object associations. This neuron was recorded from the hippocampus of a human subject in the course of clinical procedures to treat epilepsy (for details, see [57]). It responded strongly to different images of the actress Pamela Anderson and also to the letter string ‘Pamela Anderson’ (but not to other face images or letter strings). Similar results were observed for many other neurons. Such neurons explicitly encode relationships between visually dissimilar objects. We are extremely adept at learning such relationships and thus expanding our understanding of visual objects. Even someone newly exposed to American culture could quickly make the connection between Pamela Anderson’s name and face. (The numbers above each stimulus are not relevant to the results presented here.) Reproduced, with permission, from [57].

Long-term storage of object knowledge

Face perception is the extreme example of lifelong shape learning, but other aspects of object perception likewise depend on long-term memory and continual calibration. Object perception is highly inferential: the visual system learns inductive principles from the environment, taking advantage of its peculiar properties to optimize coding. For example, experience teaches that the most common lighting direction is from above, and we take advantage of that prior knowledge to derive three-dimensional shape from shading patterns. Psychophysical measurements show that even basic priors such as this can be adapted by recent experience [47].

Long-term memory may also be critical for consistent recognition of objects under changing viewing conditions, especially across viewpoint rotations [48]. For moderate rotations (up to $\sim 60^\circ$), there seems to be a computational mechanism for generalizing across views [49]. For larger rotations, the retinal appearance of a given object can change dramatically or even completely, and without prior experience generalization is poor and neural selectivity in IT is correspondingly low [50]. For these larger rotations, generalization might depend on learning temporal associations between multiple views of rotating objects [51]. Even position invariance might depend on such learned associations, because cross-position discrimination can be disrupted by prior exposure to subliminal changes in object shape during eye movements [52*].

Ultimately, learned associations constitute the knowledge that informs our understanding of objects. Pairwise associations between objects are explicitly represented in limbic structures such as perirhinal cortex [53,54], which is densely interconnected with IT. Initial formation of such associations might depend on the hippocampus, where familiar objects are strongly represented [55] and where changes in object selectivity are temporally correlated with learning episodes [56]. Object associations at the single-neuron level have even been observed in the human temporal lobe, where many neurons seem to encode the connections between names and faces (Figure 4) [57].

Concluding remarks

Transformation of object shape information in the ventral pathway is one of the most computationally complex tasks the brain performs. Correspondingly, it is one of the most difficult processes to understand. At present, we have only superficial knowledge of how object representations become compact, explicit and stable enough to support our remarkable perceptual abilities. At intermediate stages of processing, we know that transformations are geometric in nature, and include extraction of higher derivatives and multipart structure in object-based coordinates. At higher stages, transformations depend on

long-term memory: neurons are tuned in composite dimensions based on experience with familiar object categories, and object knowledge is embodied by neurons signaling learned associations. Further progress towards a unified understanding of ventral pathway function will require more quantitative analysis of geometric and mnemonic information processing.

Acknowledgements

This work was supported by the National Eye Institute and by the Pew Scholars Program in the Biomedical Sciences.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Ungerleider LG, Mishkin M: **Two cortical visual systems**. In *Analysis of Visual Behavior*. Edited by Ingle DG, Goodale MA, Mansfield RJQ. MIT Press; 1982:549-586.
2. Hung CP, Kreiman G, Poggio T, DiCarlo JJ: **Fast readout of object identity from macaque inferior temporal cortex**. *Science* 2005, **310**:863-866.
- The authors quantified information about object category and identity carried by neurons in monkey IT, using a linear classification method that might approximate the function of downstream areas such as prefrontal cortex. They showed that highly accurate discrimination can be obtained with small numbers (~ 100) of neurons analyzed across short time periods.
3. Hadjikhani N, Liu AK, Dale AM, Cavanagh P, Tootell RB: **Retinotopy and color sensitivity in human visual cortical area V8**. *Nat Neurosci* 1998, **1**:235-241.
4. Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB: **Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex**. *Proc Natl Acad Sci USA* 1995, **92**:8135-8139.
5. Kanwisher N, McDermott J, Chun MM: **The fusiform face area: a module in human extrastriate cortex specialized for face perception**. *J Neurosci* 1997, **17**:4302-4311.
6. Epstein R, Kanwisher N: **A cortical representation of the local visual environment**. *Nature* 1998, **392**:598-601.
7. Schwarzlose RF, Baker CI, Kanwisher N: **Separate face and body selectivity on the fusiform gyrus**. *J Neurosci* 2005, **25**:11055-11059.
8. Downing PE, Jiang Y, Shuman M, Kanwisher N: **A cortical area selective for visual processing of the human body**. *Science* 2001, **293**:2470-2473.
9. Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB: **Faces and objects in macaque cerebral cortex**. *Nat Neurosci* 2003, **6**:989-995.
10. Hubel DH, Wiesel TN: **Receptive fields and functional architecture of monkey striate cortex**. *J Physiol* 1968, **195**:215-243.
11. Olshausen BA, Field DJ: **Emergence of simple-cell receptive field properties by learning a sparse code for natural images 2**. *Nature* 1996, **381**:607-609.
12. Vinje WE, Gallant JL: **Sparse coding and decorrelation in primary visual cortex during natural vision**. *Science* 2000, **287**:1273-1276.
13. Pasupathy A, Connor CE: **Responses to contour features in macaque area V4**. *J Neurophysiol* 1999, **82**:2490-2502.
14. Gallant JL, Braun J, Van Essen DC: **Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex**. *Science* 1993, **259**:100-103.

15. Hegde J, Van Essen DC: **A comparative study of shape representation in macaque visual areas V2 and V4.** *Cereb Cortex* 2007, **17**:1100-1116.
16. Ito M, Komatsu H: **Representation of angles embedded within contour stimuli in area V2 of macaque monkeys.** *J Neurosci* 2004, **24**:3313-3324.
17. Brincat SL, Connor CE: **Underlying principles of visual shape selectivity in posterior inferotemporal cortex.** *Nat Neurosci* 2004, **7**:880-886.
18. Kayaert G, Biederman I, Op de Beeck HP, Vogels R: **Tuning for shape dimensions in macaque inferior temporal cortex.** *Eur J Neurosci* 2005, **22**:212-224.
19. Pasupathy A, Connor CE: **Shape representation in area V4: position-specific tuning for boundary conformation.** *J Neurophysiol* 2001, **86**:2505-2519.
20. Andrews DP, Butcher AK, Buckley BR: **Acuities for spatial arrangement in line figures: human and ideal observers compared.** *Vision Res* 1973, **13**:599-620.
21. Treisman A, Gormican S: **Feature analysis in early vision: evidence from search asymmetries.** *Psychol Rev* 1988, **95**:15-48.
22. Wilson HR, Wilkinson F, Asaad W: **Concentric orientation summation in human form vision.** *Vision Res* 1997, **37**:2325-2330.
23. Wolfe JM, Yee A, Friedman-Hill SR: **Curvature is a basic feature for visual search tasks.** *Perception* 1992, **21**:465-480.
24. Ben-Shahar O: **Visual saliency and texture segregation without feature gradient.** *Proc Natl Acad Sci USA* 2006, **103**:15704-15709.
25. Habak C, Wilkinson F, Zakher B, Wilson HR: **Curvature population coding for complex shapes in human vision.** *Vision Res* 2004, **44**:2815-2823.
26. Pasupathy A, Connor CE: **Population coding of shape in area V4.** *Nat Neurosci* 2002, **5**:1332-1338.
27. Biederman I: **Recognition-by-components: a theory of human image understanding.** *Psychol Rev* 1987, **94**:115-147.
28. Marr D, Nishihara HK: **Representation and recognition of the spatial organization of three-dimensional shapes.** *Proc R Soc Lond B Biol Sci* 1978, **200**:269-294.
29. Yamane Y, Tsunoda K, Matsumoto M, Phillips AN, Tanifuji M: **Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex.** *J Neurophysiol* 2006, **96**:3147-3156.
- The authors used a combination of optical imaging and single-unit recording in anterior IT to demonstrate clustering of cells selective for specific spatial relationships between object parts.
30. Jellema T, Perrett DI: **Neural representations of perceived bodily actions using a categorical frame of reference.** *Neuropsychologia* 2006, **44**:1535-1546.
- Neurons in the upper bank of the superior temporal sulcus exhibit complex responses to images of bodies and biological motion. This new study shows that these cells can encode body motion and posture within a reference frame that rotates with body orientation.
31. Brincat SL, Connor CE: **Dynamic shape synthesis in posterior inferotemporal cortex.** *Neuron* 2006, **49**:17-24.
32. Arguin M, Saumier D: **Conjunction and linear non-separability effects in visual shape encoding.** *Vision Res* 2000, **40**:3099-3115.
33. Wolfe JM, Bennett SC: **Preattentive object files: shapeless bundles of basic features.** *Vision Res* 1997, **37**:25-43.
34. Jiang X, Rosen E, Zeffiro T, Vanmeter J, Blanz V, Riesenhuber M: **Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques.** *Neuron* 2006, **50**:159-172.
35. McKone E, Kanwisher N, Duchaine BC: **Can generic expertise explain special processing for faces?** *Trends Cogn Sci* 2007, **11**:8-15.
36. Yovel G, Kanwisher N: **Face perception; domain specific, not process specific.** *Neuron* 2004, **44**:889-898.
37. Tsao DY, Freiwald WA, Tootell RB, Livingstone MS: **A cortical region consisting entirely of face-selective cells.** *Science* 2006, **311**:670-674.
- This study uses a powerful combination of functional imaging and microelectrode recording techniques to examine a face-selective patch in monkey IT. The patch is shown to consist almost entirely of neurons that are highly selective for faces relative to other object categories.
38. Afraz SR, Kiani R, Esteky H: **Microstimulation of inferotemporal cortex influences face categorization.** *Nature* 2006, **442**:692-695.
- The authors trained monkeys in a noise-masked face-non-face discrimination task. Microstimulation of face-selective sites in IT consistently biased responses towards the face category. This is the first study to establish a causal link between inferotemporal neural activity and a specific shape percept.
39. Rossion B, Kung CC, Tarr MJ: **Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex.** *Proc Natl Acad Sci USA* 2004, **101**:14521-14526.
40. Edelman S: *Representation and Recognition in Vision.* MIT Press; 1999.
41. Webster MA, Kaping D, Mizokami Y, Duhamel P: **Adaptation to natural facial categories.** *Nature* 2004, **428**:557-561.
- This study demonstrates that perception of face gender, ethnicity and expression is influenced by the distribution of faces observers are exposed to, both within short-term adaptation experiments and across long-term real-world experience. This shows how experience calibrates the shape boundaries used to make categorical judgments about faces.
42. Rhodes G, Brennan S, Carey S: **Identification and ratings of caricatures: implications for mental representations of faces.** *Cognit Psychol* 1987, **19**:473-497.
43. Mauro R, Kubovy M: **Caricature and face recognition.** *Mem Cognit* 1992, **20**:433-440.
44. Leopold DA, O'Toole AJ, Vetter T, Blanz V: **Prototype-referenced shape encoding revealed by high-level aftereffects.** *Nat Neurosci* 2001, **4**:89-94.
45. Loffler G, Yourganov G, Wilkinson F, Wilson HR: **fMRI evidence for the neural representation of faces.** *Nat Neurosci* 2005, **8**:1386-1390.
46. Leopold DA, Bondar IV, Giese MA: **Norm-based face encoding by single neurons in the monkey inferotemporal cortex.** *Nature* 2006, **442**:572-575.
47. Adams WJ, Graf EW, Ernst MO: **Experience can change the 'light-from-above' prior.** *Nat Neurosci* 2004, **7**:1057-1058.
48. Vetter T, Poggio T, Bulthoff HH: **The importance of symmetry and virtual views in three-dimensional object recognition.** *Curr Biol* 1994, **4**:18-23.
49. Wang G, Obama S, Yamashita W, Sugihara T, Tanaka K: **Prior experience of rotation is not required for recognizing objects seen from different angles.** *Nat Neurosci* 2005, **8**:1768-1775.
50. Freedman DJ, Riesenhuber M, Poggio T, Miller EK: **Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex.** *Cereb Cortex* 2006, **16**:1631-1644.
51. Wallis G, Bulthoff HH: **Effects of temporal association on recognition memory.** *Proc Natl Acad Sci USA* 2001, **98**:4800-4804.
52. Cox DD, Meier P, Oertel N, DiCarlo JJ: **'Breaking' position-invariant object recognition.** *Nat Neurosci* 2005, **8**:1145-1147.
- In this experiment, human observers were asked to make saccadic eye movements towards objects offset from the center of gaze. On a subset of trials, the object was surreptitiously replaced during the saccade. After many such exposures, in a discrimination task the subjects exhibited confusions specific to the objects and locations involved in the replacement events. This result suggests that position-invariant shape perception depends on learned correspondences between retinal

input patterns across changes in eye position, and that these correspondences are constantly updated on a short timescale (here less than one hour).

53. Erickson CA, Jagadeesh B, Desimone R: **Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys.** *Nat Neurosci* 2000, **3**:1143-1148.
54. Naya Y, Yoshida M, Miyashita Y: **Forward processing of long-term associative memory in monkey inferotemporal cortex.** *J Neurosci* 2003, **23**:2861-2871.
55. Yanike M, Wirth S, Suzuki WA: **Representation of well-learned information in the monkey hippocampus.** *Neuron* 2004, **42**:477-487.
56. Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA: **Single neurons in the monkey hippocampus and learning of new associations.** *Science* 2003, **300**:1578-1581.
57. Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I: **Invariant visual representation by single neurons in the human brain.** *Nature* 2005, **435**:1102-1107.